THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Allosteric Regulation of the AAA+ Superfamily Protein Reptin

Lucy Catherine Remnant

Doctor of Philosophy

The University of Edinburgh

2020

Declaration

I hereby declare that I am the author of this thesis. The work herein is entirely my own unless otherwise clearly indicated and acknowledged. I can confirm that this thesis has been submitted for the degree of Doctor of Philosophy and no part of this work has been submitted for any other degree or professional qualification.

Lucy Remnant

19/11/2020

Abstract

Reptin is a member of the AAA+ (adenosine triphosphatase associated with diverse cellular activities) protein superfamily with roles in many protein complexes and interactions which have been indicated in cancer including transcription factors and chaperon proteins. An *in silico* screen was used to identify small molecule binders of the ATP binding site of Reptin. These small molecules were then investigated for their effect on known protein-protein interactions of Reptin, specifically the Reptin-AGR2 interaction. Chemical optimisation of the original 'hit' compound lead to the development of Liddean, whose binding was seen to affect the Reptin-AGR2 interaction as well as the oligomerisation of Reptin. The affect of Liddean on Reptin was then further investigated using a combinational peptide phage library to find binding motifs, this led to the discovery of the novel Reptin interaction site on the p53 tumour suppressor protein. Hydrogen deuterium exchange was used to gain further information into the changes in conformation and oligomerisation seen with the addition of Liddean. This revealed a novel site on Reptin, forming a inter-subunit bridge, responsible for stabilisation of the Reptin homo-hexamer, known as the 'Click-motif'. Further study identified tyrosine 340 as crucial for the formation of the Reptin homo-hexamer by point mutation to alanine resulting in Reptin which is monomeric *in vitro*. These data suggest that the Reptin interactome is regulated ligand binding and the conformational changes which occur due to this.

Lay Summary

Reptin is a member of the large AAA+ (adenosine triphosphatase associated with diverse cellular activities) superfamily. Reptin is known to have interactions with several proteins important for cancer, including many involved in cell growth and cell death. A virtual screen was used to identify drugs which could bind directly to Reptin that would have an affect on its protein-protein interactions. The ‘hits’ that were identified were tested using a known complex, Reptin-AGR2, which has previously been associated with metastasis, the process in which caner spreads around the body. These investigations lead to the discovery of Liddean, a specific small molecule binder of Reptin which can alter its interaction with AGR2. To confirm that Liddean was able to cause changes in the protein-protein interactions of Reptin, a library of short peptide motifs was exposed to Reptin with and without Liddean and binding motifs were identified. These experiments identified p53, a protein known to have important roles in cancer, as another binder of Reptin. Further investigations looked at how Liddean and Reptin bound together and revealed a specific part of Reptin which is necessary for it to form a complex with itself. The precise location of the self-interaction was studied, and a single amino acid was found to be required. The data collected in this study show that the protein-protein interactions of Reptin can be controlled by the binding of small molecules to the protein.

Acknowledgements

I want to thank my Mum, my Nan and my late Grandad for supporting me through everything. You have been there whenever needed at the end of a phone call to calm me down when things are not going right and to congratulate me when things are going well, I could not have asked for more from you all, without you I would not be here writing this.

I want to thank Rebecca Rodriguez, spelling and punctuation checker and all-time best lab partner, you got me through my undergraduate degree and have been a massive rock throughout all the others.

Thank you to all members of the Hupp and Ball labs, Val Brunton and Nick Gilbert, paying particular attention to Fiona Taggart, without your guidance and fun I would not have managed to complete three years of this, for the laughs and the tears we have all been through together I thank you.

A huge thank you to Bill Earnshaw and Margarete Heck, this would not have been possible without your support and input in this thesis. You were crucial in the completion of my thesis to a level I am proud of.

Contents

Table of Figures

Glossary of terms

AAA	ATPases Associated with diverse cellular Activities superfamily
AAA+	ATPases Associated with diverse cellular Activities plus superfamily
ADP	adenosine diphospahte
ATM	ataxia-telangiectasia-mutated
ATP	adenosine triphospahte
ATPase	ATP hydrolysis
ATR	ataxia telangiectasia and Rad3-related
BSA	bovine serum albumin
CDK	cyclin dependant kinase
cDNA	complementary DNA
CRC	color-rectal carcinoma
Cryo-EM	cryogenic electron microscopy
DAPI	4’,6’-diaminido-2-[henylindole hydrochloride
DNA	deoxyribonucleic acid

DNA-PK	DNA-dependent serine/threonine protein kinase
EM	electron microscopy
EtOH	ethanol
FCS	feotal calf serum
FITC	fluorescein isothiocyanate
GTP	Guanosine-5'-triphosphate
H2AX	histone 2AX
H3	histone 3
H4	histone 4
HCC	hepatocellular carcinoma
HDX	hydrogen-deuterium exchange
kDa	kilo Daltons
MEM	minimal essestial media
mRNA	messenger RNA
PBS	phospahte buffer saline
PCR	polymerase chain reaction

PIKK	Phosphatidylinositol 3-kinase-related kinases
PLK	polo-like kinases
PVDF	polyvinylidine difluoride
RNA	ribonucleic acid
RNP	ribonucleoprotein
rRNA	ribosomal RNA
RuvBL1	RuvB like 1
RuvBL2	RuvB like 2
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser/Thr	serine/threonine
snoRNPs	small nucleolar RNP
snRNA	small nuclear RNA
Swrl1	SWR lupus locus 1
TBP	TATA-box-binding protein
TIP60	histone acetyltransferase Tip60 complex

TRITC	tetramethylrhodamine-5-(and 6)-isothiocyanate
tRNA	transfer RNA

Chapter 1 Introduction

1.1 The treatment of Cancer

Cancer is the common name for a group of diseases associated with abnormal cell growth, which has the potential to spread to other parts of the body (Pathak et al., 2018). There are more than 200 types of cancer that are diagnosed and treated in specific ways. There are two mains ways of grouping cancers. One method is grouping by are by the type of cell that the tumour has developed from, carcinoma, sarcoma, leukaemia, lymphoma and myeloma and brain and spinal cord tumours. The second is classification by localisation of the primary tumour such as breast and liver cancers, many of these classifications have existed since the 1970s (Carter and Salvik, 1974). These different types of cancer all have different 5-year survival rates, many of these survival rates have been increased in the past 20 years however some cancers still have a very short life expectancy after diagnosis. According to the WHO, in 2014 there were 8.8 million deaths due to cancer (Stewart and Wild, 2014), this shows that cancer is still a current problem for the medical and pharmacological communities.

1.1.1 Drug development in cancer

Currently, widely used therapies are chemotherapy and radiotherapy in combination with surgery. For some types of cancer, these techniques are very successful but they also come with many adverse side effects, which, in some cases cause the patient to want to stop treatment early due to the reduction in quality of life (Hoelder et al., 2012).

Small molecule therapies are now used more commonly in the clinic with many being approved in recent years. The reason for this surge in treatment type is due to the ability of small molecules to circulate the body and affect cancer cells at multiple

sites rather than focusing on the area of the primary tumour (as in many surgical approaches) and not treating metastatic sites (cancer cells which have broken away from the primary tumour and begun to proliferate), with a reduction in the side effects seen with chemotherapy and radiotherapy as they are more specific for the mutated cancer cells. However, the development and approval of new small molecular therapies is still a complex and lengthy process (Hoelder et al., 2012).

Following on from the human genome project that announced in 2000 that nearly the whole of the human genome had been sequenced (Collins and McKusick, 2001), new techniques have been developed in DNA sequencing making it both faster and more cost-effective. This means it is now possible to sequence a patient's tumour to discover the specific mutations which have led to its change from a 'normal' cell to a cancerous one. This can then help with treatment plans, which are thought to be more specific and have a greater chance of being effective (Meldrum et al., 2011). This has now been developed even further to allow for DNA sequencing of tumour cells found in the circulatory system allowing for less invasive procedures as well as earlier detection (Heitzer et al., 2015).

Proteins identified as being highly upregulated or mutated in either a specific cancer type or many cancer types are the focus of small molecule drug development (Pathak et al., 2018). There are many different families of proteins which are targeted including kinases, cyclins, and tumour suppressor proteins.

1.1.1.1 Kinases as small molecule drug targets

Protein kinases are enzymes involved in signal transduction by the transfer of phosphate groups from donor molecules, such as ATP, to specific substrates, this process is known as phosphorylation (Wang and Cole, 2014). Phosphorylation can

both activate and inhibit substrate proteins depending on the site of the phosphorylation and its relation to the rest of the protein.

Polo-like kinases (PLK) are serine/threonine kinases. The most well-known is PLK-1, a mitotic checkpoint regulator with roles in chromosome segregation, cytokinesis and centromere maturation. Overexpression of PLK-1 results in genetic instability as a result of the loss of checkpoint control. PLK-1 has been identified as a target for drug therapy due to its inhibition resulting in disruption of the cell cycle and a reduction in proliferation (Strebhardt and Ullrich, 2006). Several small molecules are at different stages of drug trials including rigosertib (Komrokji et al., 2013) which is currently in phase 3 for blood cancers, specifically myelodysplastic syndrome but could also be effective in other leukaemia types.

Aurora kinases are also serine/threonine kinases involved in the regulation of the cell cycle, specifically mitosis and cytokinesis. Aurora A promotes CDK1 activation and mitotic entry, by phosphorylating PLK-1. Aurora A also has roles in centromere maturation, spindle assembly and spindle orientation (Bischoff and Plowman, 1999). Aurora B is part of the chromosomal passenger complex where it helps to control many stages of mitosis and cytokinesis (Carmena et al., 2012). Overexpression of both kinases has been observed in several cancer types pointing to them as a possible drug target, with many investigations ongoing including trials with Alisertib which is specific for Aurora A (Friedberg et al., 2014).

1.1.1.2 Cell cycle regulators as small molecule drug targets

As cancer is a disease of misregulated cell division, cell cycle regulatory proteins are of interest when looking for small molecule drug targets. Cyclins and Cyclin dependant kinases (CDKs) are proteins which enable the cell to progress through

the cell cycle. The cell cycle and the specific Cyclins and CDKs controlling it at each step can be found in figure 1.1 (Wasserman, 2020).

CDK4 and CDK6 control the G1 phase of the cell cycle. The gene containing these proteins is often mutated in human cancers suggesting that their misregulation has a large impact on the ability of the cells to form a tumour. Some breast cancers are currently being treated with inhibitors of these proteins, such as Palbociclib, Ribociclib and Abemaciclib which block the cells in G1, halting the cell cycle before the replication of DNA in S-phase.

CDK1 is activated by cyclins B2 and B-protein. Throughout the cell cycle, phosphorylation and dephosphorylation of CDK1 are essential for tight regulation of progression from G1 to M phase. When cells are blocked using the CDK1 checkpoint, they are unable to enter M-phase. TG-02, Flavopiridol, Roscovitine, Dinaciclib, AT7519, Milciclib and RGB-286638 are all in clinical trials at various stages (Goh et al., 2012).

1.1.1.3 Tumour suppressor proteins as small molecule drug targets

There are a wide number of known tumour suppressor proteins which can prevent cancer by a variety of mechanisms.

One such protein is BRCA1, it is involved in DNA strand break by being involved and repeat and prevention. There are many cellular functions which contribute to the tumour suppression caused by BRCA1, including apoptosis, DNA repair and transcription. BRCA1 and BRCA2 act together to stabilise the human genome. Mutations in the BRCA1 gene are found to result normally in breast and ovarian cancers, several drugs are currently in trials to aid those with BRCA1/2 mutations including Valiparib, Olaparib and Rucaparib (Kaufman et al., 2009).

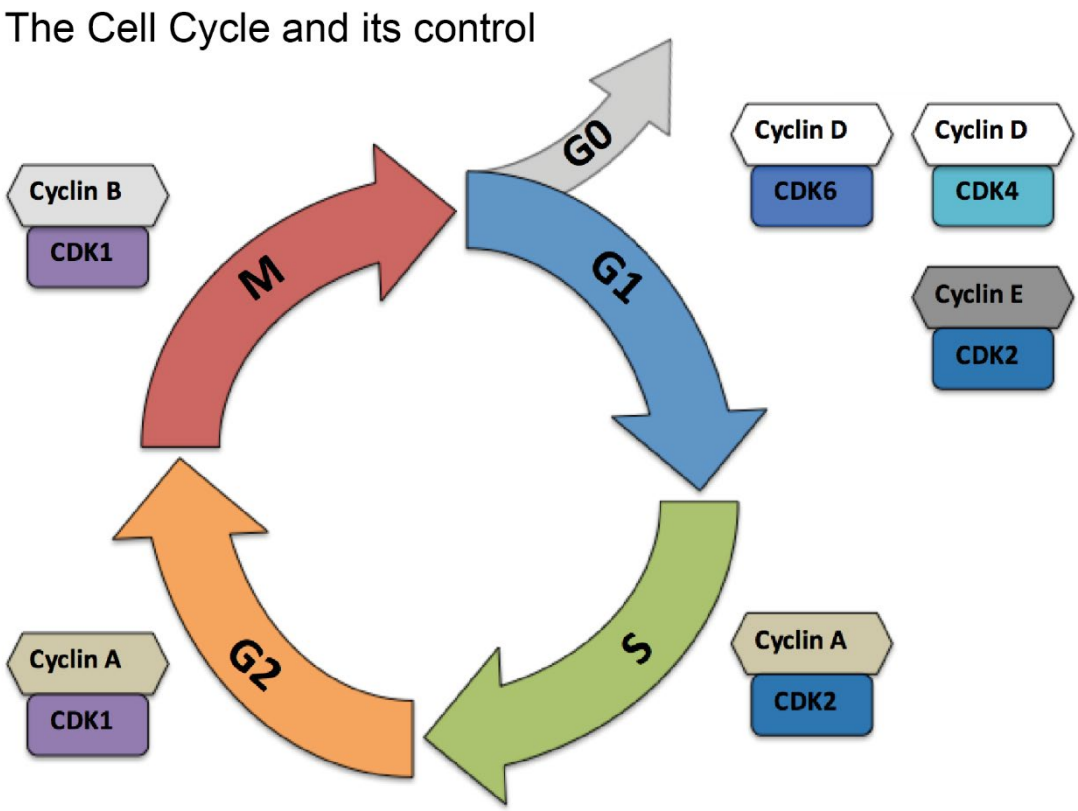


Figure 1.1 The cell cycle and its control
Cell cycle stages and the specific Cyclins and CDKs that control progression from one stage to the next. Figure adapted from (Wasserman, 2020).

P21 is a CDK1 inhibitor activated by p53. It helps to maintain cell growth and responds to DNA damage. It helps to suppress the cell cycle and DNA replication by its interaction with Proliferating cell nuclear antigen (PCNA). It also has roles in cell morphogenesis, motility, survival, apoptosis and gene transcription. HDAC inhibitors promote p21 activation resulting in inhibition of cancer cell growth. A p21 specific inhibitor is currently in development for the treatment of several cancer types and is known as Bortezomib.

Human p53 is known as the 'guardian of the genome' and has roles in controlling cell cycle checkpoints. Around 50% of human cancers have mutations in the p53 gene. Normally, p53 senses DNA damage and promotes G1 arrest to allow repair of the damage, at increased levels of damage, p53 can induce a cascade resulting in apoptosis. PRIMA-1 is currently in clinical trials as a drug targeted to restore the activity of mutant p53 proteins in cancer cells (Lewis, 2015).

1.2 The AAA+ protein superfamily

The AAA+ (ATPases associated with various cellular activities) superfamily of proteins is present within all kingdoms of life. These proteins are characterised by the presence of a conserved module of around 130 amino acids which contains Walker A and Walker B motifs, the sequence of these can be found in figure 1.2A. These motifs were first identified in 1982 by Walker and his co-workers as nucleotide-binding motifs found in several ATP binding proteins (Walker et al., 1982). These sequences were found to create a common nucleotide binding fold which enables the creating of pocket for ATP binding and hydrolysis. This superfamily of proteins is a member of the larger superfamily of ring-shaped P-loop NTPases which use their activity through the energy dependant remodelling or translocation of macromolecules, of which most are found to hydrolyse the β-g phosphate bond of a bound NTP, often ATP or GTP (Snider et al., 2008).

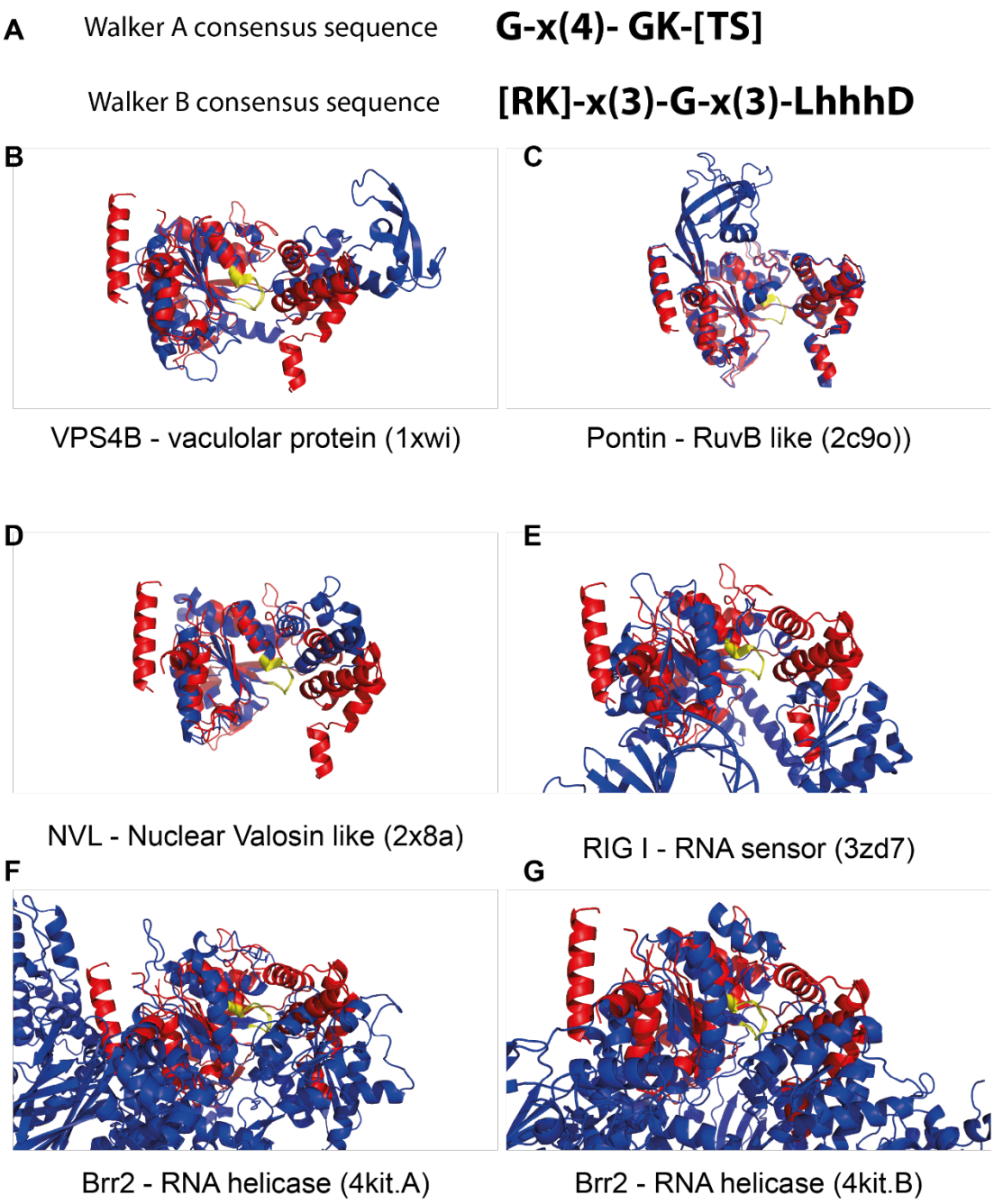


Figure 1.2 Walker A and B Motifs
A) Walker A and B motif consensus sequences B-G) Alignments of Reptin (red) with various other AAA+ superfamily proteins (blue) aligning to the Walker A motif highlighted in yellow. B) Reptin and VPS4B C) Reptin and Pontin D) Reptin and NVL E) Reptin and RIG I F) Reptin and Brr2 first site G) Reptin and Brr2 second site

AAA+ proteins are known to combine ATP hydrolysis with conformational changes in the protein which can then be transduced into a mechanical force which is then exerted on a macromolecular substrate (Hanson and Whiteheart, 2005). Often, these proteins form homohexamers which form a ring-shaped structure with a central pore, this structure helps them to carry out their function using ATP hydrolysis to induce conformational changes in the subunits which induce action upon a target substrate such as proteasomes (Vale, 2000).

As is suggested by their name, this protein family has a diverse number of functions within the different cellular compartments and have key roles in numerous essential cellular activities. Many of the AAA+ proteins also form part of large complexes with non-AAA+ proteins. Some family members can be found in figure 1.2B-G aligning their Walker A sites to the Walker A site of Reptin highlighting the ATP binding pockets of these proteins. These alignments show how the superfamily vary in the overall structure. VPS4B (Scott et al., 2005), Pontin (Matias et al., 2006), NVL (Fujiwara et al., 2011), RIG I (Kohlway et al., 2013) and Brr2 (Mozaffari-Jovin et al., 2013) are used as examples from different clades.

1.2.1 **Different clades of AAA+ proteins**

Due to their wide range of functions, more classification of the superfamily is required. This is done by dividing the family into several groups (Ammelburg et al., 2006) and clades (Iyer et al., 2004), these have been identified structurally and computationally by clustering analysis. The HEC group containing the clamp loader and DnaA/OCR clade, the extended AAA clade/group, and the PACTT group containing the helix-2 insert clade and the Lon/CLP-D2 clade, the majority of AAA+ proteins fit into one of these groupings, however, some are not able to be linked with other AAA+ proteins. The clamp loader clade proteins contain an SR[CAT] motif associate with an arginine finger, proteins found in this clade include those of the

HolB/DNAX family and the RFC family. The DnaA/OCR clade are identified by two helices of approximately equal length after strand 2 and include the DnaA and CDC6/OCR families. The AAA clade has a small helix between strand-2 and helix2 as well as an [GN]R motif associated with an arginine finger, this is the 'classical' clade and containing proteins of the FtsH, CDC48 and Proteasomal ATPase families. The Lon/CLP-D2 clade contains an extended loop after strand 2 and includes, as the name would suggest, the Lon family, the CLP family and the HslU/Clpx family. The helix-2 insert clade contains an a-b-a insert after helix-2 and the proteins in this family are from the MCM family, the chelates family, the McrB family and the dynein family.

Due to their variety, a vast number of techniques have been used to characterise several of the AAA+ proteins, with Dynein possibly being the most well characterised (Kim et al., 2012). X-ray crystallography, cryo-electron microscopy and solution NMR have been used to look at the structure of dynein alone (Kon et al., 2012) and in complex with interacting proteins (Urnavicius et al., 2018, Rao et al., 2013). The structure of the Minichromosomal Maintenance (MCM) helicase from Sulfolobus solfataricus has been examined by hydrogen-deuterium exchange to confirm the external binding sites for DNA on MCM5 (Graham et al., 2016). Due to the diversity of these proteins, a multifaceted approach to understanding their structure, function and interaction is essential.

1.2.2 **The active site of AAA+ proteins**

Due to the presence of the Walker A and B motifs, it is relatively easy to compare the structure of the different active sites of AAA+ proteins. In figure 1.2 the are several AAA+ proteins aligned to RuvBL2/Reptin. Vacuolar protein (Vps4p) is involved in the endosome (Xiao et al., 2007), nuclear vasolin like (NVL) is involved in the assembly of the telomerase holoenzyme (Her and Chung, 2012), RIG-I is

involved in RNA binding and can activate the inane immune machinery (Kohlway et al., 2013), RNA helicase (Brr2) which is a core component of the spliceosome (Mozaffari-Jovin et al., 2013) and Pontin is the homolog if Reptin and will be discussed further (Matias et al., 2015). These 6 proteins are all involved in very different processes within the cell. These were chosen due to their wide range of activities and the presence of a solved structure in the protein databank. From this is it clear that the Walker A and B sites allow for the creation of a binding pocket in all of these proteins and that the shape of this binding pocket appears to be very similar across the proteins.

1.2.3 **The Ruv proteins**

RuvB is a member of the AAA+ protein superfamily and was discovered *in E. coli* after mutations in the *ruv* gene were observed, creating both ruvA and ruvB. The *ruv* gene was named as it was discovered in *E. coli* that have a resistance to ultraviolet light in 1982 (Shurvinton and Lloyd, 1982). In the study which found the mutations, *E. coli* were exposed to low doses of ultraviolet radiation and the two mutants were found to be about 10 to 20 times more sensitive to exposure than the wildtype strain. These strains were found to form multi-nucleate filamentous cells suggesting that the ruv gene is required for septum formation in *E. coli* after periods of inhibition of DNA synthesis. In figure 1.3 the similarity between RuvA and RuvB can be found. These two proteins have been found to work together in complex and, with the addition of ATP, can reform denatured cruciform structures in supercoiled DNA. The complex is known as RuvAB and is a helicase involved in mediating Holliday junction migration (West, 1996). RuvB has been found to possess weak ATPase activity, which is stimulated by the presence of its homolog RuvA and DNA (Tsaneva et al., 1993).

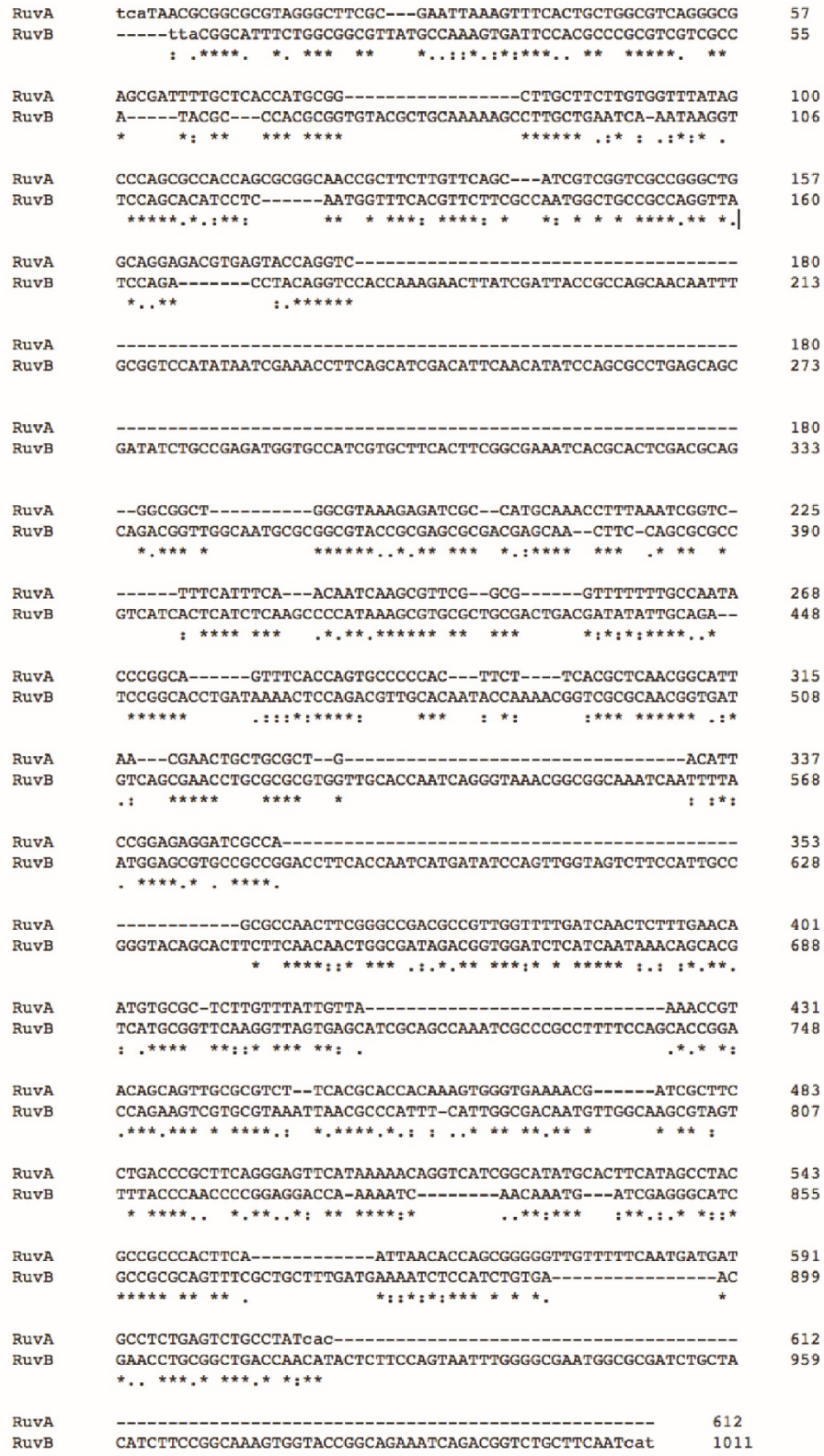


Figure 1.3 Amino acid alignments of RuvA and RuvB yeast proteins.
*Identical amino acids are identified with *, similar amino acids are highlighted with :, weakly similar amino acids are highlighted with .*

1.2.3.1 Reptin and Pontin

Reptin and Pontin are members of the AAA+ protein superfamily and were first identified as putative DNA helicases due to their homology with the bacterial protein RuvB, described above. Duplication and the evolution of the RuvB gene occurred creating these two proteins. In a similar manner to RuvB, these proteins act in a complex with one another, as RuvB did with RuvA. Both Reptin and Pontin many names. Pontin is also known as RuvBL1, Tip49a and Reptin are also known as RuvBL2, Tip49b amongst others and this is due to their wide range of functions (Matias et al., 2015). In figure 1.4 a schematic of the sequences of Reptin and Pontin is shown, highlighting the important Walker A and B motifs which classify these are AAA+ proteins as well as the sensor domains which are involved in ATP hydrolysis and/or binding. The sequences of the two proteins also show high similarity between the protein sequences with the amino acid alignment.

The oligomeric nature of these proteins both in combination and individually have been investigated. Crystal structures are complete for the heterododecamer 2XSZ (Gorynia et al., 2011) of Pontin and Reptin and the homo-hexamers of Reptin 3UK6 (Petukhov et al., 2012) and Pontin 2C9O (Matias et al., 2006). The heterododecamer is a combination of the two proteins at an equimolar ratio. The presence of these structures shows that when not in complex with each other, both proteins can form stable hexameric structures like that of half of the Reptin-Pontin dodecamer with slight variation in conformation.

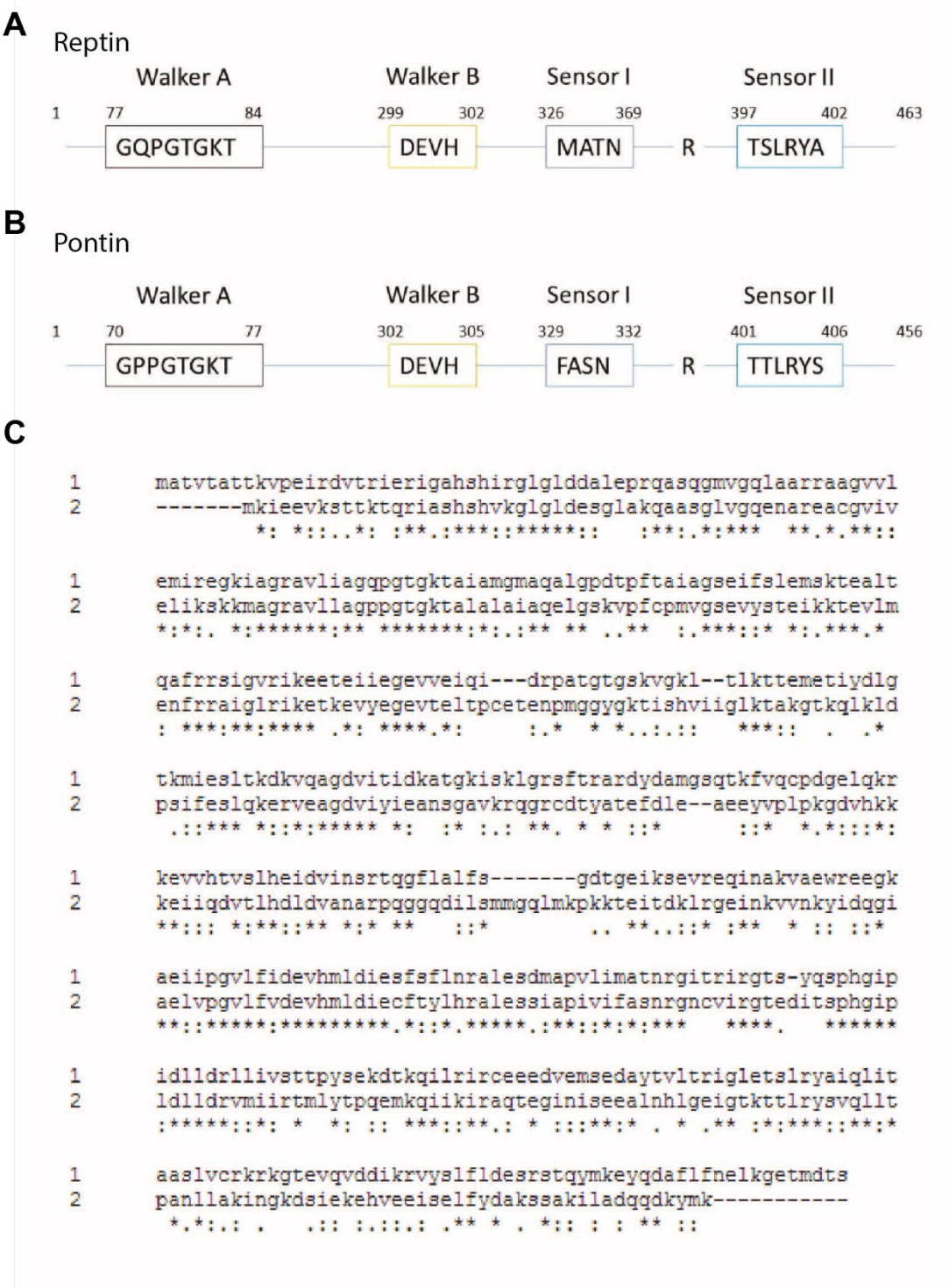


Figure 1.4 The similarities between Reptin and Pontin
*Schematic representation of A) Reptin and B) Pontin with structural elements highlighted. C) Amino acid alignment of Reptin and Pontin. Identical amino acids are identified with *, similaramino acids are highlighted with :, weakly similar amino acids are highlighted with .*

Cryo-EM has also been performed on the dodecameric complex, the data from this study shows that the complex is likely formed of two hexameric rings which interact at the head (Torreira et al., 2008). A more recent cryo-EM experiment showed that it is likely two conformations for the dodecamer, a stretched and a relaxed form (Silva-Martin et al., 2016). The compact form is more common when there is no ligand present and the percentage of stretched for increases in the presence of ADP. There is also a crystal structure available for this protein complex 2XSZ (Gorynia et al., 2011). In this crystal structure, it appears that there are two hexameric rings on top of one another with alternating subunits of Reptin and Pontin, found in figure 1.5. The similarity in the structure of the monomeric subunits of these proteins can be seen in complex in panels C and D, although variations are seen in the sequences of the two proteins, the structure of the two proteins remains very similar. Reptin can be seen in red and pontin in blue showing very small changes in their 3D alignment. The amino acid sequence and the 3D organisation of the monomeric units of Reptin and Pontin appear to show very high similarity as is expected with a gene duplication event.

The Walker A site of many AAA+ proteins is very important, in figure 1.6 we can see this motif created by aligning the Walker A sites for all human AAA+ proteins. The localisation of the site on Reptin and Pontin monomers is also visualised.

These two proteins act as chaperones and are involved in the assembly of many complexes within the cell. They are important for DNA damage repair and RNA metabolism. In many cases, Pontin and Reptin are found in the same complexes, however, they seem to have opposed functions. These proteins have also been observed to operate individually within the cell, sometimes with opposing effects such as their roles in transcriptional regulation as Pontin acts as a co-activator

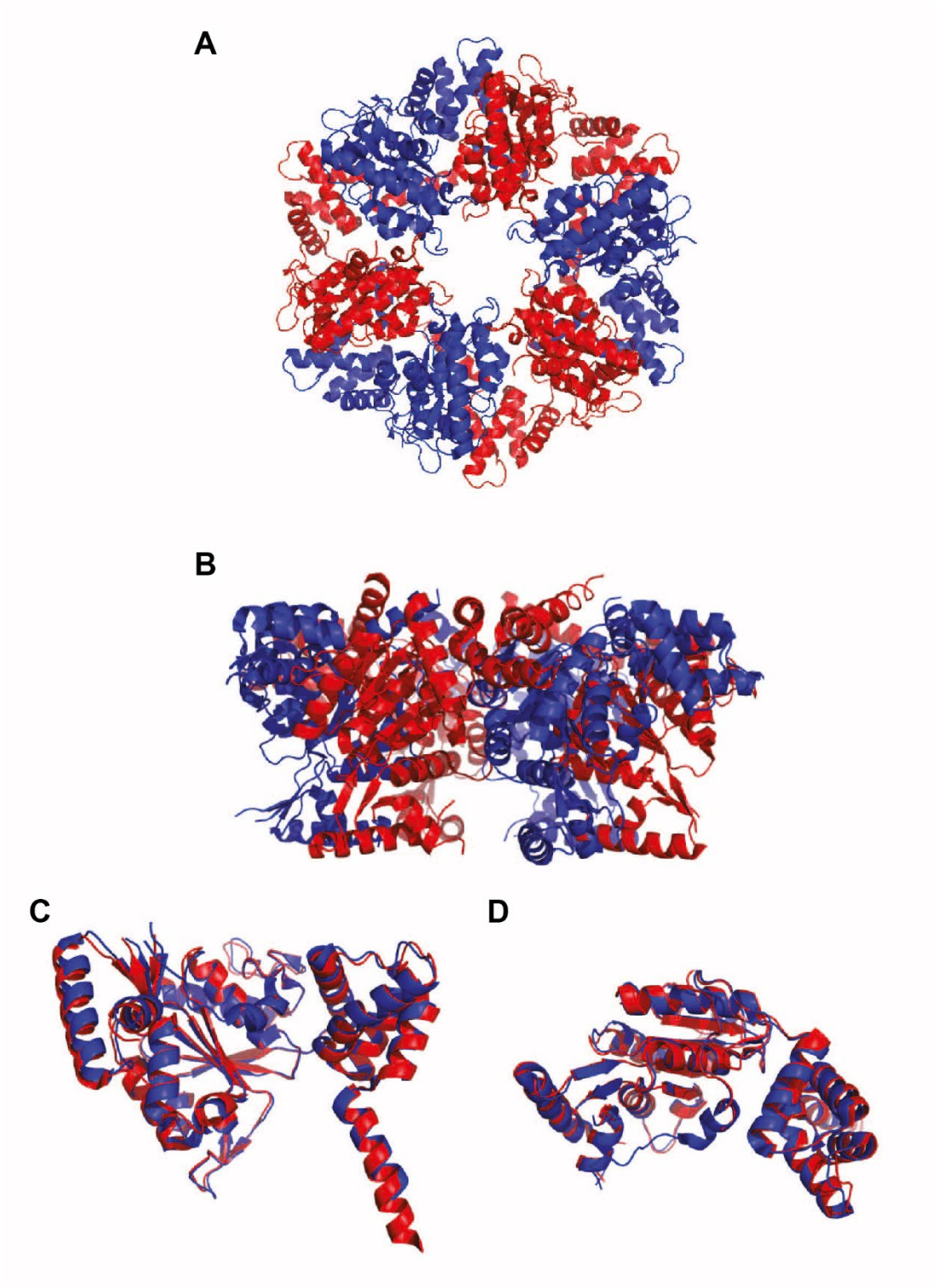


Figure 1.5 Aligning Reptin and Pontin
Reptin (red) and Pontin (blue) heterohexamer (2xsz) from the top (A) and side (B). Alignments of
single monomer of Reptin (red) and Pontin (blue) from the side (C) and top (D).

A Walker A Motif of 17 AAA+ Proteins

18.	18	1	3.00e-7	GPPGCGKT
2.	2	1	3.00e-7	GPPGCGKT
12.	12	1	1.56e-6	GPPGVGKT
6.	6	1	3.31e-6	APTGCGKT
21.	21	1	5.68e-6	GPPGTGKT
15.	15	1	5.68e-6	GPPGTGKT
13.	13	1	5.68e-6	GPPGTGKT
20.	20	1	6.52e-6	GPPGNGKT
14.	14	1	6.81e-6	GPNGSGKS
3.	3	1	8.80e-6	GPPGTGKS
10.	10	1	1.05e-5	GQVGCGKS
4.	4	1	1.05e-5	APTGAGKT
1.	1	1	1.05e-5	GQPGTGKT
11.	11	1	1.51e-5	GRTGAGKS
17.	17	1	1.82e-5	GPSGAGKS
5.	5	1	1.90e-5	APTGSGKT
16.	16	1	2.14e-5	GPAGTGKT
9.	9	1	2.42e-5	GRTGSGKS
7.	7	1	2.42e-5	GVNGVGKS
19.	19	1	2.77e-5	GPSGSGKS
8.	8	1	5.06e-5	GSTGAGKT

B

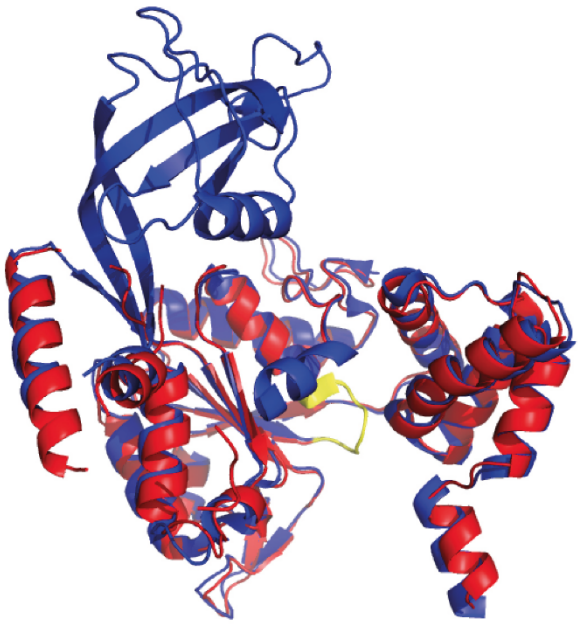


Figure 1.6 Walker A motif and alignment
A) Walker A motif as discovered using MEME suite in putting human Walker A sites.
B) Alignment of single monomer of Reptin (red) (3UK6) and Pontin (blue) (2c9o) with Walker A site in yellow.

whereas Reptin functions as a co-repressor, however, these interactions are far less characterised (Huber et al., 2008).

1.2.3.2 The role of Reptin

Reptin itself is 463 amino acids in length and around 51,000 Daltons in size, making the hexamer around 300,000 Daltons. It was discovered, in many yeast- two-hybrid assays and co-immunoprecipitation experiments throughout many model organisms (Qiu et al., 1998). This makes it a good target for biochemical analysis as there is a large difference between the oligomeric forms of Reptin making them easily distinguishable from one another. Although the Reptin homohexamer has been characterised in vitro and can be seen in the crystal structure (Petukhov et al., 2012), there is little evidence for its role in vivo. More investigation is also required to see if this homo-hexamer is present in vivo.

As previously stated, the main binding partner of Reptin is its homolog Pontin. These proteins in combination are found in many larger protein complexes within the cell including INO80, Swrl1 and TIP60 complex which are all involved in chromatin remodelling (Nano and Houry, 2013, Kim et al., 2006, Jonsson et al., 2001). They have also been observed to have roles in stimulation and repression of transcription factors such and Myc (Wood et al., 2000) and β-catenin (Kim et al., 2005). These proteins have also been observed in the assembly of ribonucleoprotein complexes like snoRNPs and telomerase (Venteicher et al., 2008, Nano and Houry, 2013). These proteins have roles in binding to heat shock proteins, such as HSP90 as part of the R2TP complex (Muñoz-Hernández et al., 2019). These interactions are summarised in figure 1.7.

The R2TP complex is made up of four proteins. This complex is highly conserved from yeast to humans. The R2TP complex was discovered in 2005 and classified as

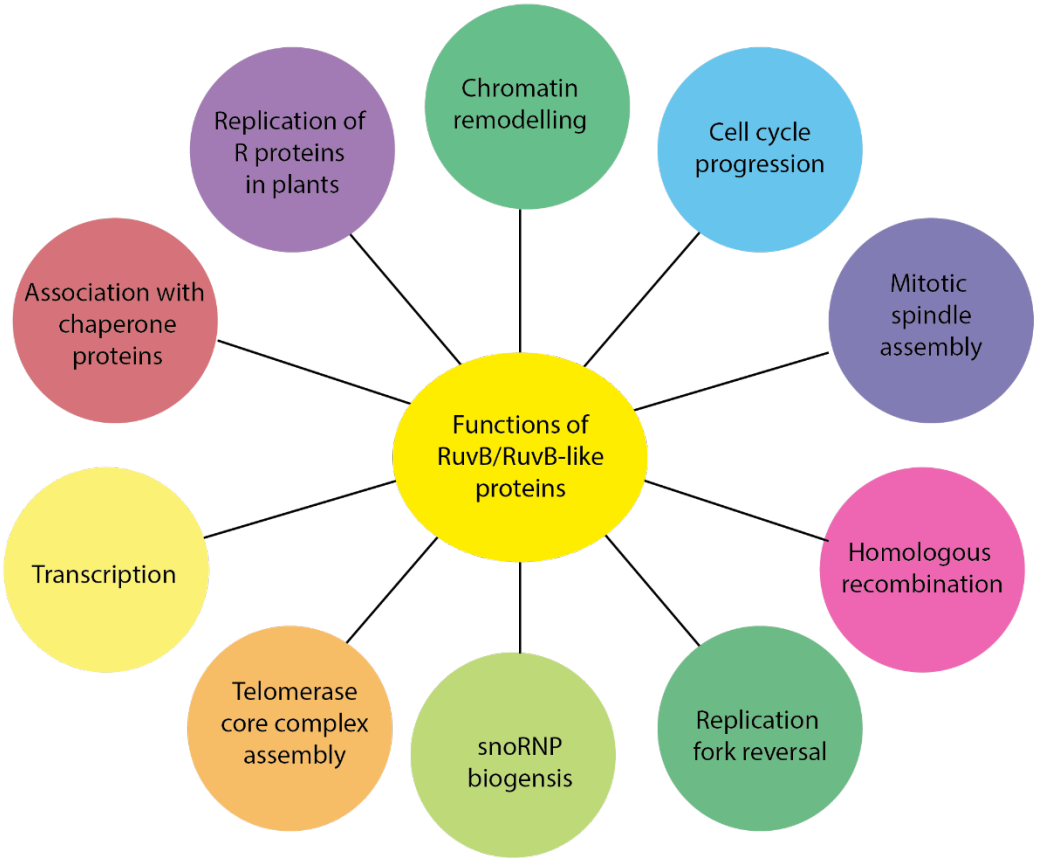


Figure 1.7 Various functions of the RuvB/RuvB-like proteins

having roles apoptosis, PIKK signalling, RNA Pol II assembly and snoRNP assembly (Kakihara and Houry, 2012). Pontin and Reptin are two of the components of the complex, and they interact with Pih1 (PIH1D1) and Tah1 (RPAP3) to form R2TP. Both Pih1 and Tah1 are Hsp90 interacting proteins, and this was thought to be the initial role of the complex. Further investigation resulted in the discovery of many cellular interactions.

The complex itself is formed by a hexameric ring of Reptin and Pontin, the Pih1-Tah1 dimer than binds to the ring via the IRD II domain of Pih1. This structural make up was discovered using a combination of cryo-electron microscopy, hydrogen-deuterium exchange, analytical ultracentrifugation, and chemical crosslinking (Rivera-Calzada et al., 2017).

TIP60 complex contains many proteins as its subunits and has a range of activities. It is characterised as tumour suppressor chromatin-remodelling complex (Ikura et al., 2000). Its activities include a well-established role in transcription, through binding and acetylation of multiple substrates including histones H2AX, H3 and H4 (Ikura et al., 2015, Sun et al., 2009, Matias et al., 2015, Sanchez-Molina et al., 2014), the Androgen Receptor (AR) (Gaughan et al., 2002) or p53 (Tang et al., 2006). The TIP60-p53 interaction is critically important for the decision between cell cycle arrest and apoptosis, a vital role of p53 with regards to cancer. TIP60 is critical for ATM acetylation and this modification precedes ATM activation. Additionally, functional human TIP60 complex is required for the recruitment of Rad51, a protein catalysing homologous recombination in the repair of DNA, to the sites of DNA damage. The presence of the Ruv proteins in the TIP60 complex is required for the assembly of a functional complex, here Reptin and Pontin act as molecular adaptors for the assembly and stability of this large multiprotein complex however the ATPase activity of these proteins is not required for this role (Jha et al., 2013).

The role of Reptin in activation or repression of transcription is mediated by the interaction of Reptin with the specific transcription factors, such as TBP, Myc, E2F1 and ATF2 (Shen et al., 2000). Reptin has also been found to bind directly to the promoters of the genes targeted by the transcription factors stated, this often involves the TIP60 complex. Reptin potentiates c-Myc mediated repression of p21 gene, hence impedes cell cycle arrest (Wood et al., 2000). Further, Reptin represses b-catenin-dependent reporters. This may indirectly point to Reptin playing a role in cellular transformation, as deregulation of β-catenin signalling, due to an activating mutation of β-catenin or its interacting partners, was observed in several cancers (Kim et al., 2005).

The important physiological function of Reptin in the chromatin remodelling complexes is linked to the DNA damage response. The DNA damage response checkpoints involve three members of the phosphatidylinositol 3-kinase-related protein kinase (PIKK) family: ATM, ATR, and DNA-dependent protein kinase catalytic subunit (DNA-PKCs) and clamp loader/polymerase clamp (RFC/PCNA)-related Rad17-RFC/9-1-1 complex (this is a complex which contains another AAA+ protein) (Izumi et al., 2010). Upon DNA damage ATM kinase autophosphorylates at Ser1981. This converts an inactive dimer into an active monomer that can bind to chromatin and phosphorylate its substrates. Activated ATM phosphorylates histones to "mark" damaged sites and this triggers recruitment of amplifying or repairing proteins, like MDC1, MRN complex, 53BP1, and BRCA1 (Izumi et al., 2012). Also, chromatin remodelling machines have to be recruited to facilitate opening and repairing of the altered chromatin structure. Interestingly, the Reptin containing chromatin remodelling complexes such as TIP60 and INO80 are also recruited to DNA damage sites.

As Reptin is incorporated in these complexes and is believed to be essential for their assembly, these findings suggest that Reptin may be involved in DNA repair pathway due to the requirement for chromatin remodelling at the DNA damage sites.

snoRNPs are RNA processing enzymes and were found to modify small nuclear RNA (snRNA), ribosomal RNA (rRNA), and tRNAs. Reptin protein was found to associate with small nucleolar RNPs (snoRNPs) in different species and be involved in their biogenesis and assembly in the process involving a chaperone multicomplex of Tah1 (Rpap3), Pih1 (Nop17), Pontin and Hsp90 (Rosenbaum et al., 2013). In addition to snoRNP, Reptin is involved in the assembly of the telomerase complex. Telomerase complex consists of the telomerase reverse transcriptase (TERT), telomerase RNA component (TERC) and the TERC-binding protein dyskerin (Her and Chung, 2012). Reptin together with Pontin interact with dyskerin and this is important for the assembly and stability of the TERC RNP. Moreover, both these proteins interact with TERT. Interestingly this PPI is regulated by the cell cycle and is increased in the S phase. The TERT-Reptin-Pontin complex displays reduced telomerase activity and precede the formation of the active telomerase complex which coincides with Reptin/Pontin dissociation. Also, the transiency of this interaction is consistent with the chaperoning activity of these proteins (Mao and Houry, 2017).

Due to this wide range of interacting proteins, the roles of Reptin include, but are not limited to, cell transformation, metastasis, response to DNA damage, apoptosis, and mitosis.

1.3 Reptin and Cancer

Reptin has been discovered to be up-regulated in some human cancers, such as prostate and hepatocellular carcinoma, making it a desirable drug

target. The interactions between Reptin and several other proteins such as Pontin (Matias et al., 2015), p53 (Tang et al., 2006), ARG2 (Maslon et al., 2010), c-Myc (Wood et al., 2000) and ARF (Xie et al., 2012), have been characterised. Many of these interactions relate to the role of Reptin in transcription and could set a starting point for the discussion on the roles of Reptin overexpression in the pathogenesis of the disease. Due to its role in both proliferative and apoptotic pathways, this protein is a favourable drug target. As there is still uncertainty of the cell signalling network surrounding this protein in the absence of its homolog Pontin, further investigation is required.

The role of Pontin and Reptin in the development of hepatocellular carcinoma (HCC) is well-established (Grigoletto et al., 2013). Not only are they both overexpressed in HCC tissues, where their overexpression was associated with poor prognosis, they both also showed stronger cytoplasmic staining in tumour cells compared to normal hepatocytes. In colorectal cancer, Reptin cooperates with Ets2 to transcriptionally regulate hTERT (Flavin et al., 2011).

The discovery of Reptin and Pontin as having roles in both HCC and colorectal cancer lead to more work being carried out on them in other cancer types. Reptin and/or Pontin have been seen to be overexpressed in oesophageal squamous cell carcinoma, gastric cancer, pancreatic ductal adenocarcinoma, renal cell carcinoma, micropapillary carcinoma, non-small cell lung cancer, small cell lung cancer, early-stage breast cancer and ductal carcinoma, ovarian cancer, acute myeloid leukaemia and lymphoma (Mao and Houry, 2017). For example, Reptin expression has also been

seen to be positively regulated by MLL-fusion proteins in acute myeloid leukaemia (Osaki et al., 2013). Mis-localisation of Reptin and/or Pontin have also been shown to be correlated with metastasis and unfavourable outcomes (Rousseau et al., 2007).

The large number of cancers which these proteins are overexpressed in suggest that they could have a role in cancer development, either alone or in partnership. The investigation into whether the contribution or regulation of tumour progression is specific to each type of cancer or can be more generalised would be of great help to understand why these proteins, which have toles in so many DNA related complexes, are found to be overexpressed in such a wide variety of cancer types.

1.4 Hypothesis

The members of the AAA+ protein superfamily have roles which are known to be highly regulated, some of which are dependent on ATP binding and ATP hydrolysis. These proteins are known to be involved in many protein complexes with both AAA+ proteins and others. The role of Reptin when combined with Pontin is well established and investigated (Mao and Houry, 2017, Rosenbaum et al., 2013) but its role alone is less clear. It is known that Reptin can form a homohexamer (Petukhov et al., 2012) but the purpose of this protein complex within cells is not understood. The work in this thesis is focused on the role of Reptin in the absence of Pontin, looking at changes in protein interactions and oligomerisation that are caused by the binding of ligands. Following on from previous work with AAA+ proteins it is likely that binding of ligands will cause a change in oligomerisation and its protein interactions.

1.5 Aims

The purpose of the investigations in this thesis is to discover more about the allosteric nature of the AAA+ protein Reptin.

We want to investigate:

- If an ATP mimetic can be created to enable tighter control of the oligomeric state of Reptin
- If the protein interactions of Reptin are dependant on ligand binding
- If Reptin can form stable interactions with other proteins in the absence of Pontin
- How the Reptin hexamer is formed and the differences between it and the Reptin-Pontin com

Chapter 2 Materials and Methods

2.1 Reagents, chemicals and plasmids

Chemicals and reagents were purchased from Sigma unless otherwise indicated. Plasmid for Reptin WT pDEST15 was a gift from Dr Magdalena Maslon, University of Edinburgh. pCDNA- 3xHA-Reptin and pCDNA-3xFLAG-Pontin were obtained from addgene. The wild type AGR2 protein was a gift from Dr Terry Gray and the wild type p53 protein was a gift from Kathrin Luise. Peptides (see table below) were purchased from Chiron Mimotopes and were synthesised with an N- terminal biotin tag and Ser-Gly-Ser-Gly spacer.

Table of peptides used in protein-peptide binding assays

Protein	Identifier	Sequence
Reptin	1	SGSGMATVTATTKVPEIRD
Reptin	2	SGSGPEIRDVTRIERIGAH
Reptin	3	SGSGRIGAHSHIRGLGLDD
Reptin	4	SGSGLGLDDALEPRQASQG
Reptin	5	SGSSQASQGMVGQLAARRA
Reptin	6	SGSGAAARRAGVVLEMIRE
Reptin	7	SGSGEMIREGKIAGRAVLI
Reptin	8	SGSGRAVLIAGQPGTGKTA
Reptin	9	SGSGTGKTAIAMGMAQALG
Reptin	10	SGSGAQALGPDTPFTAIAG
Reptin	11	SGSGTAIAGSEIFSLEMSK
Reptin	12	SGSGLEMSKTEALTQAFRR
Reptin	13	SGSGQAFRRSIGVRIKEET
Reptin	14	SGSGIKEETEIIEGEVVEI
Reptin	15	SGSGEVVEIQIDRPATGTG
Reptin	16	SGSGATGTGSKVGKLTLKT
Reptin	17	SGSGLTLKTTEMETIYDLG
Reptin	18	SGSGIYDLGTLMIESLTKD
Reptin	19	SGSGSLTKDKVQAGDVITI
Reptin	20	SGSGDVITIDKATGKISKL
Reptin	21	SGSGKISKLGRSFTRARDY
Reptin	22	SGSGRARDYDAMGSQTKFV
Reptin	23	SGSGAEWREEGKAEIIPGV
Reptin	24	SGSGIIPGVLFIDEVHMLD
Reptin	25	SGSGVHMLDIESFSFLNRA
Reptin	26	SGSGFLNRALESDMAPVLI
Reptin	27	SGSGAPVLIMATNRGITRI
Reptin	28	SGSGGITRIRGTSYQSPHG
Reptin	29	SGSGQSPHGIPIDLLDRLL
Reptin	30	SGSGLDRLLIVSTTPYSEK
Reptin	31	SGSGPYSEKDTKQILRIRC
Reptin	32	SGSGLRIRCEEEDVEMSED
Reptin	33	SGSGEMSEDAYTVLTRIGL
Reptin	34	SGSGTRIGLETSLRYAIQL
Reptin	35	SGSGYAIQLITAASLVCRK
Reptin	36	SGSGLVCRKRKGTEVQVDD
Reptin	37	SGSGVQVDDIKRVYSLFLD

Reptin	38	SGSGSLFLDESRSTQYMKE
Reptin	39	SGSGQYMKEYQDAFLFNEL
Reptin	40	SGSGLFNELKGETMDTS
p53 peptide 38	1	SGSGYFTLQIRGRERFEMF
p53 peptide 38	2	SGSGYATLQIRGRERFEMF
p53 peptide 38	3	SGSGYFALQIRGRERFEMF
p53 peptide 38	4	SGSGYFTAQIRGRERFEMF
p53 peptide 38	5	SGSGYFTLAIRGRERFEMF
p53 peptide 38	6	SGSGYFTLQARGRERFEMF
p53 peptide 38	7	SGSGYFTLQIAGRERFEMF
p53 peptide 38	8	SGSGYFTLQIRARERFEMF
p53 peptide 38	9	SGSGYFTLQIRGAERFEMF
p53 peptide 38	10	SGSGYFTLQIRGRARFEMF
p53 peptide 38	11	SGSGYFTLQIRGREAFEMF
p53 peptide 38	12	SGSGYFTLQIRGRERAEMF
p53 peptide 38	13	SGSGYFTLQIRGRERFAMF
p53 peptide 38	14	SGSGYFTLQIRGRERFEAF
p53 peptide 38	15	SGSGYFTLQIRGRERFEMA
p53 peptide 38	16	SGSGFTLQIRGRERFEMF
p53 peptide 38	17	SGSGTLQIRGRERFEMF
p53 peptide 38	18	SGSGLQIRGRERFEMF
p53 peptide 38	19	SGSGQIRGRERFEMF
p53 peptide 38	20	SGSGIRGRERFEMF
p53 peptide 38	21	SGSGYFTLQIRGRERFEM
p53 peptide 38	22	SGSGYFTLQIRGRERFE
p53 peptide 38	23	SGSGYFTLQIRGRERF
p53 peptide 38	24	SGSGYFTLQIRGRER

2.2 Equipment

A Fluroskan Ascent FL (Thermo Scientific) and Victor3 1420 Multilabel Counter (Perkin Elmer) were used to read 96-well plates. DNA concentrations were measured using a Nanodrop(R) spectrophotometer. SDS-PAGE was carried out using Bio-Rad Protean II mini-gel system. Radioactivity-containing plates were visualised with the use of Phosphoimager. X-ray films were developed using a Konica Medical Film Processor (Model SRX-101A). Sorvall RC-5C Plus, Eppendorf 5810 R and Eppendorf 5415R were used for all centrifugations. PCR was performed using the DNA Engine Dyad Peltier cycler (Bio-Rad).

2.3 Microbial techniques

2.3.1 Growing bacterial cultures

Bacterial cultures were grown in Luria-Bertani (LB) media, at 37°C, with shaking 200 rpm, and in the presence of a selective antibiotic when required, at the following final concentrations: 100 μg/ml ampicillin, 50 μg/ml kanamycin, 30 μg/ml chloramphenicol. Cultures were inoculated from a single colony or glycerol stocks, and grown in sterile flasks with a capacity of at least four times the culture volume, to allow appropriate aeration.

LB media

1 % (w/v) bacto-tryptone

0.5 % (w/v) bacto-yeast extract

1 % (w/v) NaCl

Dissolve in distilled water and autoclave at 121°C for 20 minutes.

LB agar plates were prepared using LB media with the addition of 1.5 % (w/v) bacto-agar. LB agar was first melted by heating in the microwave and then cooled to around 45°C. Subsequently, the appropriate antibiotic was added and LB agar was poured into 90 mm Petri dishes (Sterilin), and allowed to cool. The culture plates

were stored at 4°C for no longer than 2 weeks and were warmed to 37°C for 1 hour before use.

2.3.2 **Glycerol stocks**

Glycerol stocks were prepared to allow long term storage of bacteria containing plasmids or phage. 800 μl of an overnight culture was mixed with 200 μl of 80 % sterile glycerol and stored in a cryotube (Nunc), this was snap-frozen in liquid nitrogen and stored at -80°C.

2.3.3 **Preparation of competent cells**

Bacterial cells were inoculated into 5 ml of LB and incubated overnight at 37°C with shaking at 200 rpm. The overnight culture was diluted 1:200 in 100 ml of LB and incubated at 37°C until the OD600 reached 0.6. Cultures were centrifuged for 20 minutes at 4000 rcf, 4°C with the resulting pellets resuspended in 15 ml of ice-cold buffer I. This was then incubated on ice for 1 hour before centrifugation as described above. The pellet was resuspended in 4 ml of ice-cold buffer II and incubated for 15 minutes on ice. Subsequently, the cells were aliquoted (50 μl) into pre-chilled sterile Eppendorf tubes, snap-frozen in liquid nitrogen and stored at -80°C.

Buffer I

60 mM CH_3COOK

10 mM RbCl

10 mM $CaCl_2$

40 mM MgCl

15 % (v/v) Glycerol

Adjust to pH 5.8 with CH_3COOH and filter sterilise.

Buffer II

10 mM MOPS

10 mM RbCl

75 mM CaCl2

15 % (v/v) Glycerol

Adjust to pH 6.5 with NaOH and filter sterilise.

2.3.4 Transformation of bacterial cells

100 ng of plasmid DNA was added to 50 μl of freshly thawed competent cells and incubated on ice for 30 minutes. The cells were heat-shocked at 42°C for 1 minute and placed straight on ice for 2 minutes to cool. 500 μl of LB was added, and the culture was incubated at 37°C for 45 minutes, with shaking 200 rpm. 100 μl was then plated into LB-agar plate containing appropriate selective antibiotic and incubated at 37°C overnight.

2.4 Molecular Biology Techniques

2.4.1 Amplification, purification, and quantification of plasmid DNA

A single bacterial colony was picked from an LB-agar plate and inoculated into 5 ml of LB broth containing selective antibiotic when required and grown for several hours at 37°C with shaking 200 rpm. The starter culture was diluted into 500 ml on LB containing selective antibiotic, if required, and grown overnight at 37°C with shaking 200 rpm. The culture was then centrifuged at 4°C for 15 minutes at 6000 rcf. Plasmid DNA was extracted from pellet using plasmid Maxi kit (Qiagen) according to the manufacturer's instructions. If lower yield was required starter cultured was subjected to DNA Mini kit (Qiagen). Plasmid DNA was eluted in nuclease-free water and stored at -20°C. The concentration of obtained DNA was measured using Nanodrop ND-1000 (absorbance at 260 nm).

2.4.2 Agarose gel electrophoresis of DNA

Agarose gel electrophoresis was used to separate and analyse DNA. Agarose gels were prepared by adding electrophoresis-grade agarose (Invitrogen)to 1x TAE

buffer to a final concentration of 1 % and melted by heating in the microwave. The agarose solution was then cooled to around 60oC before the addition of SybrSafe DNA gel stain (Thermo Fisher) to a concentration of 1x. The solution was then poured into a casting tank and allowed to sit at room temperature before being submerged in 1x TAE buffer. DNA samples were mixed with 6x DNA loading buffer. Once samples were loaded alongside either Quick-load 100 bp DNA ladder (NEB) and/ or 1 kb DNA ladder (NEB) the gel was run at 100 V for approximately 1 hour with the bands being visualised under UV transilluminator Syngene (Genesnap).

1x TAE

40mM Tris

1mM EDTA pH 8.0 Adjust pH to 8.0

6x DNA loading buffer

0.25 % bromophenol blue

0.25 % xylene cyanol FF

15 % Ficoll

2.4.3 **DNA sequencing**

All Sanger DNA sequencing was carried out by Source Bioscience (LifeSciences). Plasmids were sequenced using stock primers from Source Bioscience.

Next-generation sequencing was carried out by Otogenetics.

2.4.4 **Cloning**

2.4.4.1 Gateway cloning

Sequences of interest were cloned into the Gateway system (Invitrogen) following the manufacture's protocol. Initially, an insert flanked by the attB recombination sites were generated. This was followed by the production of an entry clone using the BP reaction. Finally, a destination clone was created using the LR reaction.

2.4.4.1.1 Producing attB-PCR Products

An insert flanked by the attB recombination sites were generated, this was performed by the creation of primers (Sigma) containing the N or C termini of the protein of interest and the corresponding attB site. The PCR reaction was carried out as is described in the PCR methodology. The PCR product was purified using gel electrophoresis and gel extraction.

2.4.4.1.2 BP and LR reactions to create entry and destination clones

BP reaction was carried out as manufactures instructions to create an entry clone using pDONR221. LR reactions were then carried out as manufactures instructions to insert Reptin and pontin into desired destination vectors.

2.4.4.2 Conventional cloning

Conventional cloning is performed using restriction enzymes (RE) and is carried out in three steps, amplification, digestion and ligation.

2.4.4.2.1 PCR amplification of the desired insert, with the addition of the required restriction sites.

Suitable plasmid DNA for the insert was used as a template. Primers were designed that incorporated different RE sites into the 3' and 5' end. Nucleotide bases are added alongside the RE sites to allow for efficient binding. These nucleotides are random and are manipulated to get good G/C content and a suitable melting temperature. In it important to note that that primers used must result in an insert which will be in frame with the destination vector.

PCR reaction mix

25 μl 2x Pfu master mix

5 μl Band doctor (Rovalab)

5 ng template DNA

1 μl forward primer (20 μM stock)

1 μl reverse primer (20 μM stock)

Nuclease free water up to 50 μl

Thermal cycling conditions

Incubate at 95°C for 2 minutes

Incubate at 95°C for 20 seconds

Incubate at 68°C for 40 seconds

Incubate at 72°C for 1 minute

Cycle to step 2 for 30 cycles

Incubate at 72°C for 1 minute

Hold at 4°C forever

2.4.4.2.2 Restriction enzyme double digests of insert and vector.

All restriction digests were carried out using restriction enzymes (RE) and buffers from New England Biolabs and digest conditions recommended by the supplier.

Insert digest

40 μl PCR product

5 μl NEB buffer

0.5 μl BSA (10 mg/ml)

1 μl Enzyme 1

1 μl Enzyme 2

Nuclease free water up to 50 μl

Vector digest

5 μg destination vector

5 μl enzyme buffer

0.5 μl BSA (10 mg/ml)

1 μl Enzyme 1

1 μl Enzyme 2

Nuclease free water up to 50 μl

Double digests were incubated at 37°C for 60 minutes as recommended. If inactivation of the enzymes is possible, the RE was inactivated by incubating at 65°C for 10 minutes. Following the double digestion, the digest mix was loaded onto a 1 % agarose gel. The gel was viewed under UV light and the bands corresponding to digested insert and vector cut out and purified using the Qiagen Gel extraction kit. Purified DNA was eluted in 50 μl nuclease-free water.

2.4.4.2.3 Ligation of digested vector and insert.

Ligation of vector and insert was carried out using T4 DNA ligase (Promega), following the protocol from the supplier. 100 ng vector was used and the amount of insert required was calculated using the following formula: Insert(ng) = (Vector(ng) x Insert size(kb)) x Molar ratio of insert Vector size(kb) vector ratio of insert to vector was always tested by ligation. 1:1, 2:1 and 3:1 molar

Ligation reactions

1 μl ligase buffer (10x) 100 ng vector

X ng insert

1 μl T4 DNA ligase

Nuclease free water to 10 μl water

Ligation reactions were carried out at room temperature for 1 hour and overnight.

2.5 μl of the mix was transformed into DH5α competent cells and plated onto LB agar plates containing antibiotic matching that of the destination vector. Colonies were selected and plasmid DNA obtained using the Qiagen Mini-prep kit.

2.4.4.3 Site-directed mutagenesis

Site-directed mutagenesis was carried out by preparing primers specific to the mutation of interest and then performing a PCR reaction. The resulting solution was exposed to Dpn1 before transformation into DH5 αcompetent cells.

Reptin D299N mutant

forward 5′-GAGTGCTGTTCATCAACGAGGTCCACATGC-3′

reverse primer 5′-GCATGTGGACCTCGTTGATGAACAGCACTC-3′

Reptin Y340A mutant

forward 5′- GAATCCGGGGCACCAGCGCCCAGAGCCCTCACGGCA-3′

reverse 5′- GCCGTGAGGGCTCTGGGCGCTGGTGCCCCGGATTC-3′

PCR reaction mix

25 μl 2x Pfu master mix

5 μl Band doctor (Rovalab)

5 ng template DNA

1 μl forward primer (20 μM stock)

1 μl reverse primer (20 μM stock)

Nuclease free water up to 50 μl

Thermal cycling conditions

Incubate at 95 °C for 2 minutes

Incubate at 95 °C for 1 minute

Incubate at 60 °C for 1 minute

Incubate at 72 °C for 5 minutes

Cycle to step 2 for 19 cycles

Incubate at 72 °C for 5 minutes

Hold at 4 °C forever

Once PCR was completed, 1μl Dpn1 (20000 U/ml, NEB) used to digest methylated DNA. The incubation was carried out at 37°C for 1 hour. Dpn1 was inactivated after this by incubation at 65oC for 10 minutes.

2.5 μl of the mix was transformed into DH5α competent cells and plated onto LB agar plates containing antibiotic matching that of the destination vector. Colonies were selected and plasmid DNA obtained using the Qiagen Mini-prep kit.

2.5 Biochemical techniques

2.5.1 SDS-PAGE for protein separation

Polyacrylamide gels were prepared as described by Laemmli, using the recipes listed below and the Biorad Mini-PROTEAN 3 Cell. The appropriate separating gel was cast and overlaid with isopropanol. The isopropanol acts in two ways, it levels out the top of the separating gel and prevents oxygen from reaching the gel, allowing the even polymerisation of the acrylamide. Once the separating gel had polymerised the isopropanol was removed and the stacking gel cast. Samples to be loaded were first mixed with sample buffer (2x) in a 1:1 ratio and heated for 3-5 minutes at 90 °C. Prestained protein markers (Fermentas) were loaded and used as size markers. Gels were run at 150 V for 1 hour in running buffer (1x) and terminated once the dye front had reached the end of the gel.

10 % Resolving gel

30 % acrylamide mix 10 % (v/v)

1.5 M TRIS (pH 8.8) 0.39 M

10 % (w/v) SDS 0.1 % (v/v)

10 % (w/v) APS 0.1 % (v/v)

TEMED 0.04 % (v//v)

H2O to final volume

(Other percentage resolving gels, 8 %, 12 % and 15 % were assembled in the same manner adjusting the amount of acrylamide mix and water.)

Stacking gel

30 % acrylamide mix 5 % (v/v)

1 M TRIS (pH 6.8) 0.13 M

10 % (w/v) SDS 0.1 % (v/v)

10 % (w/v) APS 0.1 % (v/v)

TEMED 0.1 % (v//v)

H2O to final volume

Sample buffer

300mM Tris (pH 6.8)

5 % (w/v) SDS

25 % (v/v) glycerol

400mM DTT

bromophenol blue to desired colour

Running Buffer

192 mM glycine

25 mM Tris

0.1 % (w/v) SDS

2.5.2 **Coomassie staining**

After SDS-PAGE, to detect all proteins in the gels, coomassie staining was performed. Gels were submerged in Coomassie blue stain for 20 minutes and then destained to the desired colour by submerging in destain for 10 minutes before removal and addition of clean destain for the rest of the incubation (2 hours-

overnight). Once destained, the gels were rinsed in water and dried using a heated vacuum gel drier (Gel master model 1426, Welche Rietschle Thomas).

Stain

50 % (v/v) methanol

10 % (v/v) glacial acetic acid

0.2 % (w/v) coomassie brilliant blue R-250

Destain

7.5 % (v/v) methanol

10 % (v/v) glacial acetic acid

2.5.3 **Western blotting**

For analysis of specific proteins, western blotting was performed. Proteins were transferred from SDS-PAGE gels and onto 0.2 μM nitrocellulose membrane (Protran). The transfer was carried out in tanks containing transfer buffer and an ice pack, to prevent overheating, at 100 V for 1 hour or 30 mAmps overnight. Transfer apparatus supplied by Biorad.

Post transfer the membrane was washed in Phosphate Buffered Saline (PBS) containing 0.1 % (v/v) TWEEN-20 (PBST), 3 washes for 5 minutes. The membrane was blocked with blocking buffer (3 % (w/v) (skimmed milk powder in PBST) for 30 minutes. The membrane was incubated with the primary antibody in blocking buffer for 1 hour at room temperature, or overnight at 4 °C, and then washed with PBST, 3 washes for 5 minutes. The membrane was incubated with horseradish peroxidase-conjugated secondary antibody (1:1000, Dako) in blocking buffer for 1 hour at room temperature and washed as above.

Antibody signal was detected using enhanced chemiluminescence (ECL). Blots were overlaid with a fresh mix of ECL 1 and ECL 2 (1:1) for 1 minute, blotted to dry

and exposed to Hyperfilm ECL (Amersham) for the desired period. The film was developed using a Konica medical film processor (model SRX-101A).

Transfer Buffer

192 mM glycine

25 mM TRIS

20 % methanol

ECL 1

100mM Tris (pH 8.5)

2.5 mM luminol

0.4 mM p-Coumaric acid

ECL 2

100mM Tris (pH 8.5)

0.02 % hydrogen peroxide

2.5.4 **Stripping nitrocellulose blots**

When probing with multiple antibodies was required, blots were stripped by incubation with stripping buffer for 30 minutes at room temperature with gentle agitation. Blots then underwent protocol as above including washing, blocking and probing.

Stripping buffer

62.5 mM TRIS (pH 6.8) 2 % (w/v) SDS

0.6 % (v/v) BME

2.6 Cell culture

2.6.1 Cell lines and media

All cell lines were incubated at 37ºC with 5 % CO_2 (except A375 which require 10 % CO_2) in a humidified incubator. Media (Gibco) was supplemented with 10 % (v/v) FBS and 1 % (v/v) penicillin/streptomycin (Gibco).

2.6.2 Cell number and viability

To assess cell number and viability, cells were counted using a haemocytometer. First cells were diluted 1:5 with Trypan blue solution to allow differentiation between live and dead cells. To calculate the number of viable cells/ml the average cell count from each of the sets of 16 corner squares is taken and then multiplied by 10,000.

To correct for the dilution with trypan blue, multiply by 5. This gives the number of viable cells/ml.

2.6.3 Sub-culturing of cells

Cells were maintained in sterile 10 cm diameter dishes and cultured to 70-80% confluence. When the cells reached this point, culture media was removed, and the cell monolayer was washed twice with 10 ml of sterile PBS. 2 ml of 0.05% Trypsin/EDTA was added to the dish and incubated at 37ºC until the cells detached from the plastic surface. Next, 8 ml of fresh culture media was added, and the cell suspension was added to a 15 ml falcon tube. After appropriate dilution, cells were seeded into a new cell culture plate, as required.

2.6.4 Freezing and recovery of cells

At 90-100% confluence, cells were trypsinised from flasks for 5 minutes. 10 ml new media was added before centrifugation. Centrifugation of the 15 ml falcon occurred at 200 rcf for 5 minutes and the supernatant was removed. Cells were resuspended in 3 ml of freezing media. The cell was aliquoted into cryotubes (Nunc) and stored in

Nalgene TM Cryo 1ºC freezing container at -80ºC overnight before permanent storage in liquid nitrogen.

To thaw and recover cells from liquid nitrogen, a cryovial containing the cells was rapidly thawed at 37°C in a water bath. The cells were transferred into a 10 cm culture dish containing 10 ml of fresh culture media. The next day, media was discarded, and new media was put in its place. The cells were then incubated at 37°C until confluent.

2.6.5 Transient transfection of DNA

Transfection of DNA was carried out at around 50 % confluence. All transfections were carried out using Attractene (Qiagen). The amounts of DNA transfected were normalised using appropriate empty vector DNA.

For 6 well transfection, the appropriate amount of DNA was incubated with 4.5 μl Attractene in a total volume of 100 μl, made up with serum and antibiotic-free media for 20 minutes at room temperature. During this incubation 1900 μl of fresh media was added to each well of the 6-well plate. After incubation, the transfection reaction is added to the wells and incubated at 37°C for 24 or 48 hours or as is sown in figure legends. Cells were either treated or harvested as described in figure legends.

2.6.6 Drug and small molecule treatment

For these experiments, cells were treated with drugs and small molecules following the transfection protocol above.

2.6.7 Harvesting cells

Cells were placed on ice to chill and the culture media was discarded. Cells were then washed three times with ice-cold PBS (2 ml for 6-well plate and 10 ml for 10 cm plate). Cells were scraped into 1 ml of ice-cold PBS and transferred into a microfuge

tube and centrifuged for 5 minutes at 3000 rpm. The supernatant was discarded, and the cell pellets were snap-frozen in liquid nitrogen and stored at -80°C.

2.6.8 **Cell lysis**

Cells were lysed using Urea lysis buffer, between 2 and 4 volumes (depending on the size of cell pellet). Cells were resuspended in lysis buffer by pipetting and incubated on ice for 30 minutes. The cells were centrifuged for 15 minutes at maximum speed. The supernatant was transferred to a fresh tube and protein concentration was calculated using a Bradford assay.

Urea lysis buffer

8 M Urea

50 mM Tris (pH 8) 1 % (w/v) DTT

2 % (w/v) SDS

1 x cOmpleteTM, Mini, EDTA-free Protease Inhibitor Cocktail (Roche)

2.6.9 **Protein quantification**

Protein concentration was determined using the Bradford assay use Bradford's reagent (Bio-rad), following manufacturer's instructions. Absorbance at 595 nm was measured using the Victor3 plate reader with BSA concentration standards.

2.7 Protein expression and purification from *E. coli*

2.7.1 **Protein expression in *E. coli***

Wild type or mutant Reptin and Pontin expression vectors were transformed into BL21-AI cells. A single colony was inoculated into 200 ml LB containing 100 μg/ml ampicillin, and incubated overnight at 37°C with shaking at 200 rpm. 20 ml of starter culture was diluted into 1000 ml of LB containing 100 μg/ml ampicillin and incubated at 37°C, shaking until the OD600 reached 0.6. Protein expression was induced with the addition of 0.2 % (w/v) arabinose, this was incubated at 30°C for 3 hours with shaking at 200 rpm. Cultures were centrifuged at 6000 rcf for 15 minutes at 4°C.

Resulting pellets were resuspended in 30 ml of storage buffer and snap-frozen for storage in -80°C.

Storage buffer

10 % (w/v) sucrose

50 mM HEPES pH 7.5

2.7.2 **Purification of GST-tagged protein**

The cell pellet in storage buffer was defrosted in warm water for 5 minutes before being transferred into ice for the rest of its defrosting. The components of the lysis buffer were added to the storage buffer to create the lysis buffer. Cell pellet in lysis buffer was incubated on ice for 30 minutes and then sonicated 3 times for 15 seconds at an amplitude of 10 with 10-second incubations on ice between each sonication. The lysate was then centrifuged for 15 minutes at 4°C at 13000 rpm. The supernatant was then transferred to a tube containing 500 μl of glutathione-sepharose 4B beads (Amersham GE) (washed 3 times in PBS to remove storage solution) and incubated for 1.5 hours at 4°C rotating. The lysate and beads were centrifuged for 5 minutes at 4000 rpm and the supernatant removed. Next, 5 ml of high salt wash buffer was used to add beads onto a 5 ml disposable column (MoBiTec) and left to empty by gravity. Beads were washed twice with 5 ml of high salt wash buffer with triton and then twice with high salt wash buffer (without triton). Low salt washes were then performed twice with triton, twice without triton. Next, the beads were washed with 5 ml elution buffer before the column was closed and 500 μl of elution buffer was added to the column. Finally, 30 μl of PreScission protease (GE Healthcare) was added to the beads and incubated overnight rotating at 4°C. In the morning, this was centrifuged at 2000 rpm for 5 minutes and the supernatant containing cleaved protein was collected and transferred to the new tube before being stored at 4°C for short term storage or aliquoted into 100 μl and stored at -20°C.

Lysis buffer

10 % Sucrose (already present)

50 mM HEPES pH 7.5 (already present)

400 mM NaCl

0.1% Triton-X 100

1mM DTT

1mM benzamidine

0.5 mg/ml lysozyme

1x cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail

2.8 Biochemical Assays

2.8.1 In vitro peptide binding assay

Streptavidin beads were saturated with the biotinylated peptide of interested. The beads were first incubated with peptide diluted in Pierce Protein-Free Blocking Buffer (Thermo Scientific) rotating for 1 hour at room temperature. The beads were then washed to remove any residual non-binding peptide with PBST (3x 1ml) Whole-cell lysate or purified protein was then incubated with the beads, rotating at room temperature for 1 hour. Beads were washed with PBST (5x 1ml) before the proteins were eluted into SDS-PAGE sample buffer by boiling for 10 minutes 95°C. Eluted proteins were subjected to SDS-PAGE for analysis.

2.8.2 Peptide-protein binding assay

Streptavidin (1 μg per well) was coated onto a CostarTM white 96 well plate (Fisher) in water (50 μl) overnight at 37 °C. The following day wells were washed with PBST (5x 200 μl, 0.1 % (v/v) TWEEN-20) and then incubated with biotin tagged peptide, to saturate the streptavidin (500 ng per well), in water (50 μl). The wells were washed as above and blocked with either Pierce Protein-Free Blocking Buffer (200 μl, Thermo Scientific) or blocking buffer ((PBS (200 μl) containing 3 % (w/v) BSA (BSA-PBS)), for 1 hour at room temperature, with shaking. The wells were then washed

with PBST as described above. A titration of the protein of interest was added to wells (the wells preincubated with peptide) in Pierce Protein-Free Blocking Buffer or blocking buffer ((PBS (200 μl) containing 3 % (w/v) BSA (BSA-PBS)) (50μl) and incubated for 1 hour at room temperature, with shaking. The wells were then washed as described above and incubated with primary antibody diluted in Pierce Protein-Free Blocking Buffer or blocking buffer ((PBS (200 μl) containing 3 % (w/v) BSA (BSA-PBS)) (50μl) for 1 hour at room temperature, with shaking. The wells were washed as above then incubated with HRP-conjugated secondary antibody diluted in Pierce Protein- Free Blocking Buffer or blocking buffer ((PBS (200 μl) containing 3 % (w/v) BSA (BSA-PBS)) (50 μl) for 1 hour at room temperature, with shaking. The wells were washed for a final time, as described above, and peptide-protein binding was detected by electrochemical luminescence (50 μl ECL per well) and quantified using a luminometer (Labsystems Fluroscan Ascent FL).

2.8.3 **Protein-protein binding assay**

Protein 1 (100 ng per well) was coated onto a CostarTM white 96 well plate (Fisher) in 0.1 M $NaHCO^3$ (50 µl) overnight at 4°C. The wells were washed with PBST (5x 200 μl, 0.1 % (v/v) TWEEN-20) before the addition of blocking buffer ((PBS (200 µl) containing 3 % (w/v) BSA), for 1 hour at room temperature, with shaking. The wells were then washed with PBST as described above. A titration of protein 2 was added to wells (the wells preincubated with protein 1) in BSA-PBS (50 μl) and incubated for 1 hour at room temperature, with shaking. The wells were then washed as described above and incubated with primary antibody (to protein 2) diluted in 3% BSA-PBS (50 μl) for 1 hour at room temperature, with shaking. The wells were washed as above then incubated with HRP-conjugated secondary antibody diluted in 3% BSA-PBS (50 μl) for 1 hour at room temperature, with shaking. The wells were washed for a final time, as described above, and protein-protein binding was detected by

electrochemical luminescence (50 μl ECL per well) and quantified using a luminometer (Labsystems Fluroscan Ascent FL).

2.8.4 **ATPase assay**

ATPase assay was carried out in Mass spec grade H_2O using ATPase Activity Assay Kit (Colorimetric) (Biovision) following manufactures protocol. Purified proteins were added directly or incubated with each other overnight to allow for oligomerisation to occur. Detection was carried out at OD 650 nm.

2.8.5 **ATP filter binding assay**

Reptin, wildtype or mutant, is incubated with ATP-gamma$_{32}$P in solution (PBS). The solutions are then filtered through a disc to capture the protein. If the ATP is binding to the protein it will be captured with the protein on the disc. These discs are then placed on a phosphor screen to expose the amount of radioactivity in each sample. The phosphor screen is then imaged using a Phosphoimager. This creates a digital image, like a western blot, which can be used for quantitation.

2.8.6 **Crosslinking of protein using glutaraldehyde**

Reptin WT was incubated with and without 100μM Liddean for 1 hour. To perform the crosslinking, increasing concentrations of glutaraldehyde were added and incubated for 15 minutes. The protein complexes found were then analysed using SDS PAGE and Western blotting.

2.8.7 **Peptide phage display**

Peptide phage display was performed using a commercial library phage expressing different 7mer or 12mer peptides on their surface. This is exposed to the protein of interest and the peptides which bind are identified using next-generation sequencing.

Round 1 This experiment is set up in duplicate on two separate plates to analyse the data when comparing fast washes and slow washes. The first steps are simultaneous.

The purified protein (100 ng per well) was coated onto a CostarTM white 96 well plate (Fisher) in 0.1 M $NaHCO_3$ (50 μl) overnight at 4°C. The following day wells were washed with Tris-buffered saline containing (3x, 0.1 % (v/v) TWEEN-20 (TBST) and then blocked with Tris-buffered saline (TBS, 200 μl) containing 3 % (w/v) BSA. The phage (Ph.D-12 phage display library, New England Biolabs) was added to wells (4 μl in 96 μl TBST), to those wells containing the protein. For the fast wash data, the wells of one plate were washed with TBST (10x 200 μl). The phage was eluted with 0.2 M glycine-HCl pH2.2 (100 μl) containing BSA (1 mg/mg) for 15 minutes at room temperature with rocking. This step must not be allowed to progress past 15 minutes. The eluted phage was transferred to a fresh Eppendorf and 1M Tris-HCl (15 μl) was added to neutralise the phage solution. This neutralisation must be carried out immediately. The eluted phage was then stored at 4°C. For the slow data wells were washed with TBST (6x 200 μl), 5 minutes per wash. Following the washes, the phage was eluted and neutralised as before. An overnight was set up with ER2738 cells (New England Biolabs) in LB (10 ml) containing tetracycline (1 μg/μl) at 37°C, 220 rpm. The overnight culture was diluted 1:200 in LB (20 ml) and grown for 1 hour at 37°C, 220 rpm. It was then inoculated with the eluted phage (100 μl) and incubated for 4.5 hours at 37°C, 220 rpm. The culture was centrifuged for 10 minutes, 12000g, 4°C and the supernatant transferred to a fresh tube and centrifuged for 10 minutes, 12000g, 4°C. The top 80 % of supernatant was transferred to a fresh tube and 20 % (w/v) PEG/2.5 M NaCl was added (1/6 of the supernatant volume) to precipitate the phage, the phage was precipitated overnight at 4°C. In preparation of the second round the following day, purified protein (100 ng per well) was coated onto a CostarTM white 96 well plate

(Fisher) in 0.1 M NaHCO3 (50 μl) overnight at 4°C. As above the plates were set up in duplicate. The following day the PEG solution was spun for 15 minutes, 12000g, 4°C and the supernatant discarded, the tube was left upside down to drain for 10 minutes. The pellet was resuspended in TBS 200 μl, transferred to an Eppendorf and centrifuged for 5 minutes, 13000 rpm, 4°C. The supernatant was transferred to a new tube, and the amplified phage from round one was stored at 4°C.

Round 2 was carried out following the method for round one, except, instead of adding phage library to the wells, amplified phage (10 μl in 100 μl total volume in TBST) was added at this point in the method.

PCR step. The eluted and amplified phage were subjected to PCR before being run on agarose gels before next-generation sequencing by Otogenetics.

PCR mix

5 μl 5 M betaine

10 μl 5x herculase II buffer 5 μl 132 mM trehalose

2.5 μl 10 mM dNTP

1 μl 100 mM reverse primer

1 μl herculase II enzyme

1 μl forward primer

10 μl phage

Nuclease free water to 50 μl

Thermal cycling conditions

Incubate at 95 °C for 60 seconds

Incubate at 95 °C for 15 seconds

Incubate at 55 °C for 20 seconds

Incubate at 70 °C for 60 seconds

Cycle to step 2 for 30 cycles

Incubate at 70 °C for 210 seconds

Hold at 4 °C forever

2.8.8 **Antibody phage display**

Antibody phage display was performed using an in lab-made library of phage expressing different nanobodies on their surface. This is exposed to the protein of interest and the nanobodies which bind are then amplified and identified using sequencing.

Round 1 This experiment is set up in duplicate on two separate plates to analyse the data when comparing fast washes and slow washes. The first steps are simultaneous.

The purified protein (100 ng per well) was coated onto a CostarTM white 96 well plate (Fisher) in 0.1 M NaHCO3 (50 μl) overnight at 4 °C. The following day wells were washed with Tris-buffered saline containing (3x, 0.1 % (v/v) TWEEN-20 (TBST) and then blocked with Tris-buffered saline (TBS, 200 μl) containing 3 % (w/v) BSA. The phage was added to wells (4 μl in 96 μl TBST). For the fast wash data, the wells of one plate were washed with TBST (10x 200 μl). The phage was eluted with 0.2 M glycine- HCl pH2.2 (100 μl) containing BSA (1 mg/mg) for 15 minutes at room temperature with rocking. This step must not be allowed to progress past 15 minutes. The eluted phage was transferred to a fresh Eppendorf and 1M Tris-HCl (15 μl) was added to neutralise the phage solution. This neutralisation must be carried out immediately. The eluted phage was then stored at 4 °C. For the slow data wells were washed with TBST (6x 200 μl), 5 minutes per wash. Following the washes, the phage was eluted and neutralised as before. An overnight was set up with ER2738 cells (New England Biolabs) in LB (10 ml) containing tetracycline (1 μg/μl) at 37°C, 220 rpm. The overnight culture was diluted

1:200 in LB (20 ml) and grown for 1 hour at 37°C, 220 rpm. It was then inoculated with the eluted phage (100 μl) and incubated for 4.5 hours at 37°C, 220 rpm. The culture was centrifuged for 10 minutes, 12000g, 4°C and the supernatant transferred to a fresh tube and centrifuged for 10 minutes, 12000g, 4°C. The top 80 % of supernatant was transferred to a fresh tube and 20 % (w/v) PEG/2.5 M NaCl was added (1/6 of the supernatant volume) to precipitate the phage, the phage was precipitated overnight at 4°C. In preparation of the second round the following day, purified protein (100 ng per well) was coated onto a CostarTM white 96 well plate (Fisher) in 0.1 M NaHCO3 (50 μl) overnight at 4°C. As above the plates were set up in duplicate. The following day the PEG solution was spun for 15 minutes, 12000g, 4°C and the supernatant discarded, the tube was left upside down to drain for 10 minutes. The pellet was resuspended in TBS 200 μl, transferred to an Eppendorf and centrifuged for 5 minutes, 13000 rpm, 4°C. The supernatant was transferred to a new tube, and the amplified phage from round one was stored at 4°C.

Subsequent rounds were carried out following the method for round one, except, instead of adding phage library to the wells, amplified phage (10 μl in 100 μl total volume in TBST) from the previous round was added at this point in the method.

Nanobody pools were tested against the protein of interest by coating purified protein (100 ng per well) onto a CostarTM white 96 well plate (Fisher) in 0.1 M $NaHCO_3$ (50 μl) overnight at 4°C. The amplified phage was then incubated in the well for 1 hour (concentrated and dilution 1:10, 1:100 used). Phage was detected using an anti-phage antibody to allow for determination of the efficiency of the nanobody pool.

2.9 Biophysical Assays

2.9.1 Gel filtration

Reptin, WT and mutants, were incubated with and without ligands in 100μl HEPES buffer (pH 7.6) containing PBS. Samples were injected onto a 25 ml Superose-12 gel filtration column in the same buffer and fractions were collected at a flow rate of 0.5 ml/min. In the ligand-bound state, the column was pre-incubated in the same buffer with the addition of ligand. Gel filtration markers were used to calibrate the column.

50μl was taken from each of the fractions eluted from the column and bound to an ELISA plate before identification with anti-Reptin antibody as with standard protein-protein interaction protocol.

2.9.2 Hydrogen-deuterium exchange

Reptin (6 μM) was prepared in H_2O (control) and D_2O for set time intervals (1, 2, 5, 10, 30, 60 and 180 minutes) before the exchange was terminated by quenching with Trifluoroacetic acid.

Reptin was prepared in the presence of Liddean and incubated at room temperature (60 minutes) to allow for complex formation. The samples were then prepared for exchange as for Reptin.

The samples were measured on the Orbitrap Elite mass spectrometer connected to HD robot and liquid chromatography. HD exchange analysis was done in MS mode in orbitrap analyser with 120 000 resolution. Two different fragmentation methods were applied, collision-induced dissociation (CID) and higher-energy collisional dissociation (HCD) and peptide mapping were measured in MSMS mode with data dependant analysis (TOP3).

Chapter 3 The discovery and characterization of a first-in-class drug lead that binds Reptin

Some of the work in this chapter was published in 2015 in Chemical Science as a collaboration between the laboratories of Prof. Ted Hupp and Prof. Nicholas Westwood (Healy et al., 2015). Other work in the chapter was published in 2019 in Journal of Proteomics as a collaboration between the laboratories of Prof. Ted Hupp, Prof. Borek Vojtesek and Dr Douglas Houston (Remnant et al., 2019).

3.1 Introduction

3.1.1 Reptin and Pontin, members of the AAA+ protein superfamily

The AAA+ (ATPases associated with various cellular activities) protein superfamily is present in all kingdoms of life (Snider et al., 2008). The members of this protein superfamily exist as oligomers and are particularly interesting due to their ligand binding ability and allosteric regulation of protein function (Stinson et al., 2013). Reptin and Pontin are paralogs and members of this protein superfamily, they are usually found in complex with each other forming a heterododecamer. These proteins have high sequence identity and similarity (Nano and Houry, 2013) and are known to be involved in many cellular complexes, a large number of which many are involved in interacting with DNA. The heterohexameric complex of Reptin and Pontin was crystallised in the presence of ligand in 2011 by the Carrondo lab with a resolution of 3A (Gorynia et al., 2011). This structure shows alternating subunits of Pontin and Reptin forming two heterohexameric rings which lie on top of one another in a head-head formation.

In contrast to the well-characterised interactions of Reptin in the presence of Pontin, there is little information on the function and interactions of Reptin alone. The crystal structure of a Pontin hexamer was published in 2006 by the Carrondo lab (Matias et al., 2006), this was followed by the structure of the heterohexamer, as previously mentioned. Finally, a crystal structure of Reptin was first published in 2012 by Grigoriev lab (Petukhov et al., 2012). This shows that Reptin and Pontin can form stable complexes without each other in vitro, however, both require the presence of ligand. Previous research in the lab has shown that Reptin binds to Anterior-Gradient-2 (AGR2) (Maslon et al., 2010) in the absence of Pontin and other ligands, this lead to the belief that the Reptin homohexamer must have protein-protein interactions independent of Pontin.

Although Reptin is a member of the AAA+ protein family which are known to bind ATP, this is a highly transient interaction, Reptin alone has a very low affinity for ATP. It has been observed previously in the AAA+ protein family that conformational changes occur when a protein binds to its ligand or when the ligands are enzymatically broken-down releasing energy. One of the most well-characterised members of the AAA+ family is dynein which uses its AAA+ domain to provide the energy required for its role as a motor protein. The protein uses ATP driven conformational changes in the AAA+ ring to induce movement in the adjacent motor domain (Gleave et al., 2014).

3.1.2 **Mutations in AAA+ proteins**

In previous studies, following common mutations used to study AAA+ proteins (Hanson and Whiteheart, 2005), two mutant versions of Reptin were characterised, K83A which is a mutation in the Walker A binding site resulting in a loss-of-function mutant, and D299N which is a mutation in the Walker B site resulting in a loss of ATP hydrolysis. Further studies showed that contrary to its loss of ATP hydrolysis,

the D299N mutant protein has a higher affinity for ATP that the wildtype. In the same experiment, the K83A mutant showed a reduction in ATP binding compared to the wildtype (Maslon et al., 2010).

3.1.3 **Known Reptin protein-protein interactions**

Reptin was identified as a possible cancer therapeutic target due to its binding to Anterior-Gradient2 (AGR2) and its role in the interaction between AGR2 and p53. The binding interaction between AGR2 and Reptin is thought to have a role as part of a signalling complex involved in the development of pro-metastatic cancer (Maslon et al., 2010). This could be related to the role of Reptin in complexes such as TIP60 and its interaction with c-Myc. AGR2 has a wide range of cellular functions, including cell migration, cellular transformation and metastasis (Zhang et al., 2005, Reiner et al., 2009) but it is its role as in inhibitor of p53 which is most interesting (Pohler et al., 2004). As a therapeutic target, it was found that there were significant associations between levels of AGR2 expression in a tumour and either oestrogen receptor-α positivity or tamoxifen resistance (Fritzsche et al., 2007, Hrstka et al., 2010). AGR2 is found to be overexpressed in many cancer types, including oesophagus, pancreas, breast, prostate, and lung (Fritzsche et al., 2006, Zhang et al., 2005, Reiner et al., 2009, Liu et al., 2005). P53 is the most well-studied tumour suppressor protein with roles in transcriptional activation and DNA binding (Hupp et al., 1992, Raycroft et al., 1990, Pietenpol et al., 1994). P53 responds to cellular stress resulting in the induction of cell cycle arrest and apoptosis or senescence (Fridman and Lowe, 2003). Many cancers, especially those that are hereditary, are a result of mutations in the p53 gene (Perri et al., 2016). It is the role of Reptin binding to AGR2 and the effect of this interaction in pro-metastatic cancer, in the absence of Pontin, which leads to an investigation into ligands which bind Reptin.

3.1.4 The importance of protein-protein interactions in therapeutics

Understanding the network of protein-protein interactions (PPIs) is crucial in the development of novel therapeutics but the discovery of new PPIs is still a major challenge scientifically (Jubb et al., 2012). The creation of a comprehensive list of interactors for a protein of interest allows a greater understanding of how to manipulate specific cellular pathways, whilst attempting to reduce off-target effects within the cell. Using these complex networks to identify novel targets for diagnostic and therapeutic purposes remains an unexploited avenue in drug discovery (Healy et al., 2015).

Here computational biology, chemical synthesis and biochemistry were used to develop small molecule binders of the Reptin protein. The resulting ligands were then tested with a range of biochemical and cell biology approaches to characterise the changes seen in protein-protein interactions and the wider cell environment.

3.2 Aims

The experiments in this chapter aim to develop a small molecule ATP mimetic which can bind specifically to Reptin. This ATP mimetic should result in changes of binding of Reptin to its known binders in the same manner as ATP (Maslon et al., 2010).

We want to investigate:

- If an ATP mimetic can be created to enable tighter control of the allosteric nature of Reptin
- If the binding of this ATP mimetic results in changes of the protein interactions of in the same way as the binding of ATP
- If this ATP mimetic causes Reptin to change its oligomeric state
- If this small molecule can be used within cells and if there are any resulting phenotypes

3.3 The creation of Liddean

An in-silico screen was carried out using a library of commercially available small molecules to identify those which might bind to the Walker A site of Reptin. This method takes the binding site of a protein and combines it with a library of small molecules to see the likelihood of interaction occurring. The rigid-body docking program LIDAEUS (Taylor et al., 2008) was used to dock a virtual conformer library of 4.4 million compounds. The results were ranked based on their LIDAEUS score and the top 49971 compounds were tested again using Vina and Autodock, this work was carried out by Douglas Houston, the University of Edinburgh (flowchart of methodology can be seen in figure 3.1A). This led to the selection of 30 small molecules which had been predicted to bind to Reptin, these small molecules were all commercially available. To analyse their effect on Reptin, a peptide-binding assay was carried out against the AGR2 peptide, FVLLNLVY, which has previously been described to bind Reptin (Maslon et al., 2010). This characterisation was performed as previous work in the laboratory has shown that the presence of ATP had a direct effect on the binding interaction of Reptin and AGR2 (Maslon et al., 2010). The hits were ranked equally by a significant increase or decrease in binding of the protein to the peptide compared to the control, Reptin alone. Compound 1 (figure 3.1B) was identified as inducing the greatest response and was ranked highest for further investigation. Compound 1 contains a biphenyl substituent that is predicted to sit deep in the Walker A pocket where the adenine of ADP/ATP binds and a pyridine–oxazolo ring system which is predicted to extend out of the pocket (Healy et al., 2015).

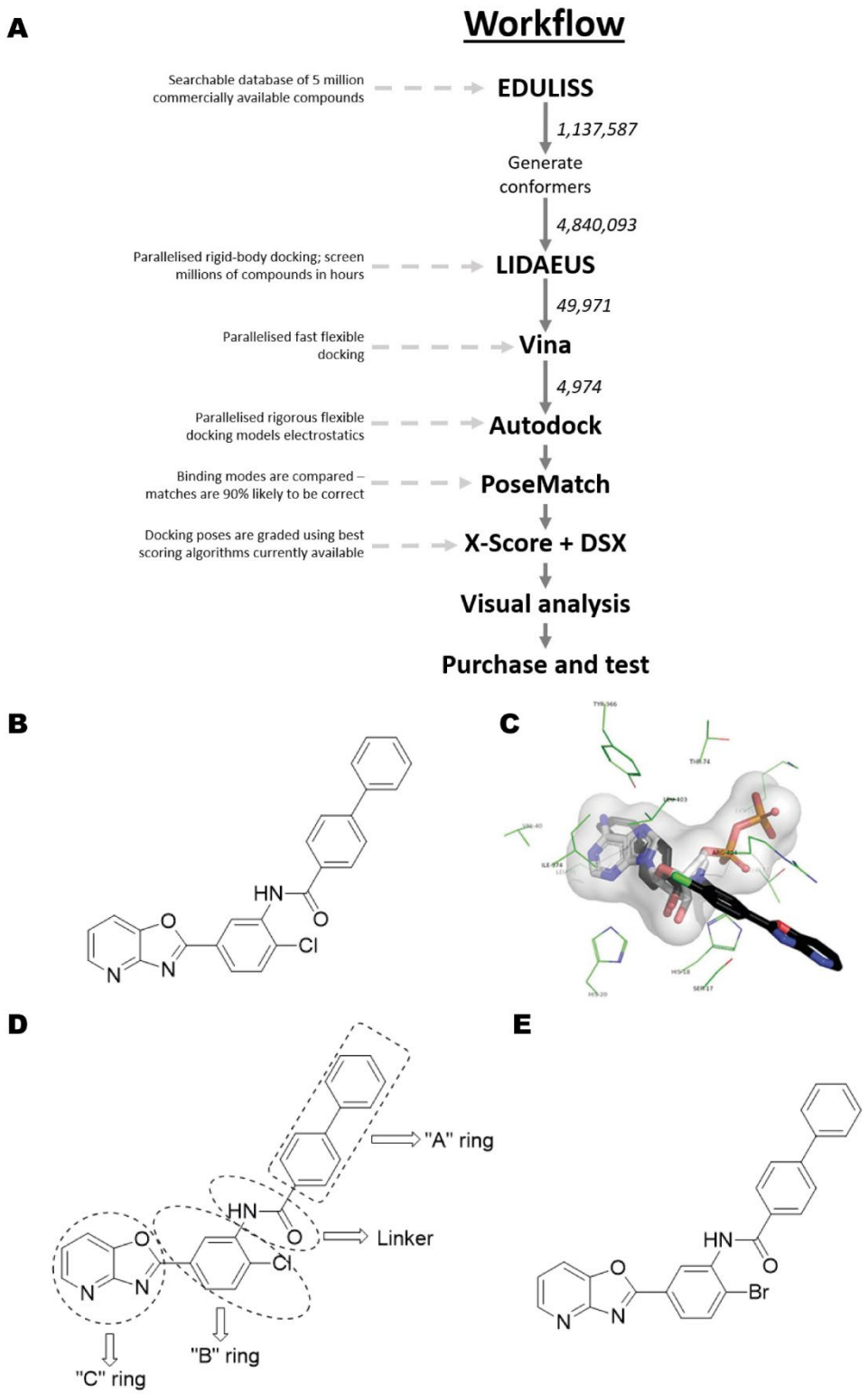


Figure 3.1 Creation of small molecule Liddean using a combinatorial approach. *A) Flowchart showing process used to combine data from the surface potential of proteins with a library of conformers of commercially available small molecules to create a list of likely binding candidates. B) Structure of 'Compound 1' C) Comparison between the binding contacts of Liddean (stick) and ATP (space fill and stick) created in the active site of Reptin. D) Locations of modification sites of Compound 1 used in development of Liddean. E) Structure of 'Liddean' (Healy et al., 2015)*

A comparison of the binding contacts in the pocket shared between ATP/ADP and Compound 1 can be seen in figure 3.1C (Autodock software). The interactions deep within the binding pocket appear to be the same for Compound 1 and ATP/ADP, however, the orientation of tail of the small molecules as they exit the binding pocket are likely to have very different conformations.

A structure-activity-relationship study (SAR) was then carried out on Compound 1 by Alan Healy, The University of St Andrews. The sites for modification can be seen in figure 3.1D showing sites of the" A"," B" and" C" rings as well as the linker region. All of these sites were modified to create a library of Compound 1 derivatives. Modified analogues of Compound 1 were either synthesised (Alan Healy) or purchased (if available) for in vitro analysis using protein-protein and protein-peptide assays.

3.3.1 **Expression of Reptin**

The DNA sequence for the Reptin protein was cloned into pDEST15 using the Gateway cloning methodology (figure 3.2A) (Katzen, 2007). The vector containing the original Reptin sequence was a gift from a previous lab member, Magdalena Maslon. Primers containing either attR1, a Precission protease cleavage site and the N-terminal nucleotides of Reptin or the C- terminal nucleotides of Reptin and the attR2 site were designed and then synthesised by Sigma. These were used to copy the Reptin sequence into a Gateway entry vector pDNR221 using the LR reaction. The sequence of the entry plasmids was then checked using Sanger sequencing by Source Bioscience against the predicted sequence of the insert, combining the cleavage site and the cDNA sequence of Reptin. Once the sequence of the entry vector was confirmed, the BP reaction was carried out using the pDNR221 containing the Reptin sequence and the pDEST15 vector. The pDEST15 was selected due to it containing GST upstream of the attR1 site so that both the GST

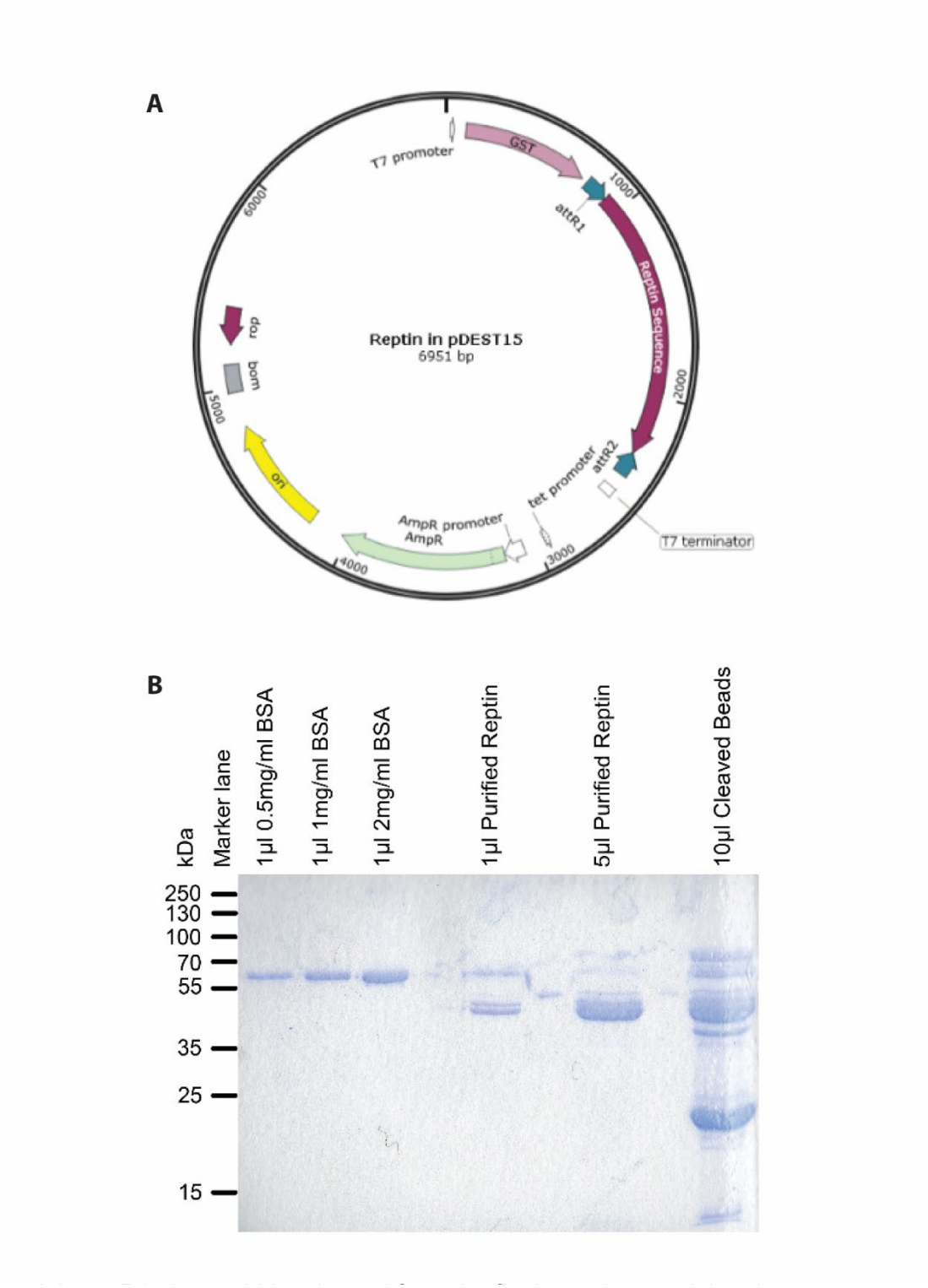


Figure 3.2 Expressing Reptin in bacteria
A) Map of the expression vector created to enable to expression and purification of human Reptin in E. coli. B) Purification of Reptin protein from bacteria showing various concentrations of BSA as standard to aid in quantification of the concentration of and purity of purified Reptin, the proteins removed directly from the beads are also shown.

expressed.

The new pDEST15 plasmid containing the Reptin sequence, from now on known as pDEST15 Reptin, was transformed into BL21-AI cells (Saida et al., 2006). These are a modified form of *E. coli* specifically for expression of toxic proteins which require tight regulation, expression of a protein with these cells requires induction with the sugar arabinose. Once transfected, the cells were grown at 37ºC overnight before being diluted the following morning and grown until the cells reached an OD of 0.8-1. The cultures were moved to 25ºC and protein expression was induced with the addition of arabinose. After induction of protein expression for 3 hours the cultures were pelleted ready for purification of the expressed protein.

For purification of the protein, the BL21-AI cells were lysed, the supernatant collected, and the expressed fusion protein was captured using Glutathione beads. The beads were washed repeatedly to remove contaminant host cell proteins bound to GST-Reptin. Finally, Precission protease is added to the beads and left to incubate overnight, the protease itself also has a GST tag meaning it is removed from the supernatant during this step. The concentration of the protein expressed is measured using the combination of a Bradford assay and a coomassie stained gel containing BSA standards, this combination allows correct quantification of the protein concentration and checks the purity of the expressed protein. An example of coomassie stained gel can be found in figure 3.2B. It is also visible in the gel that the cleaved beads still have a large amount of Reptin protein which is not eluted but has been cleaved from the GST. A smaller band is visible below 25kDa which is likely to be a combination of the GST itself and the protease used to cleave as they are both similar in weight. This method was further optimised to allow recovery of more of the purified protein by increasing both the volume of elution buffer and amount of Precission protease used to cleave the fusion protein.

3.3.2 **Testing Compound 1 analogues in vitro**

3.3.2.1 Changes in protein and peptide binding with the addition of ligands

A combination of peptide-protein and protein-protein binding assays were used to look at the effect different small molecules had on the binding affinity of Reptin to known binding proteins and itself. To keep the experimentation as comprehensive as possible, the analogues were given numbers. This allowed testing to occur 'blind' whilst still retaining the ability to look at the specific modifications which have created the greatest effect.

3.3.2.1.1 Protein-protein assays

In protein-protein assays, a polystyrene plate is coated with protein 1 (solid phase) by incubation overnight in a mildly denaturing buffer (0.1 M $NaHCO^3$), blocking solution is added for 1 hour (3% BSA in PBST), protein 2 is then added to the well in blocking solution. A primary antibody against protein 2 is used to detect the protein. A secondary antibody raised against the primary antibody (swine anti-rabbit or rabbit anti-mouse) conjugated to HRP (horseradish peroxidase) is then added. The amount of protein was quantified using the ECL reaction measuring the luminescence in relative light units. A schematic of these experiments can be found in figure 3.3A.

Reptin binding to p53 and AGR2 was used to look at the changes in protein-protein interactions. The interaction between Reptin and AGR2 peptide had been used to study the effects of the original 30 ligands, the next step was looking at protein-protein interactions. Figures 3.3C and 3.3D show representative assay results, observing the effect small molecule binders on these interactions. The binding of Reptin in solution phase is shown to change when Reptin has been incubated with the small molecule library for 1 hour before combining with the protein in the solid phase, either AGR2 or p53.

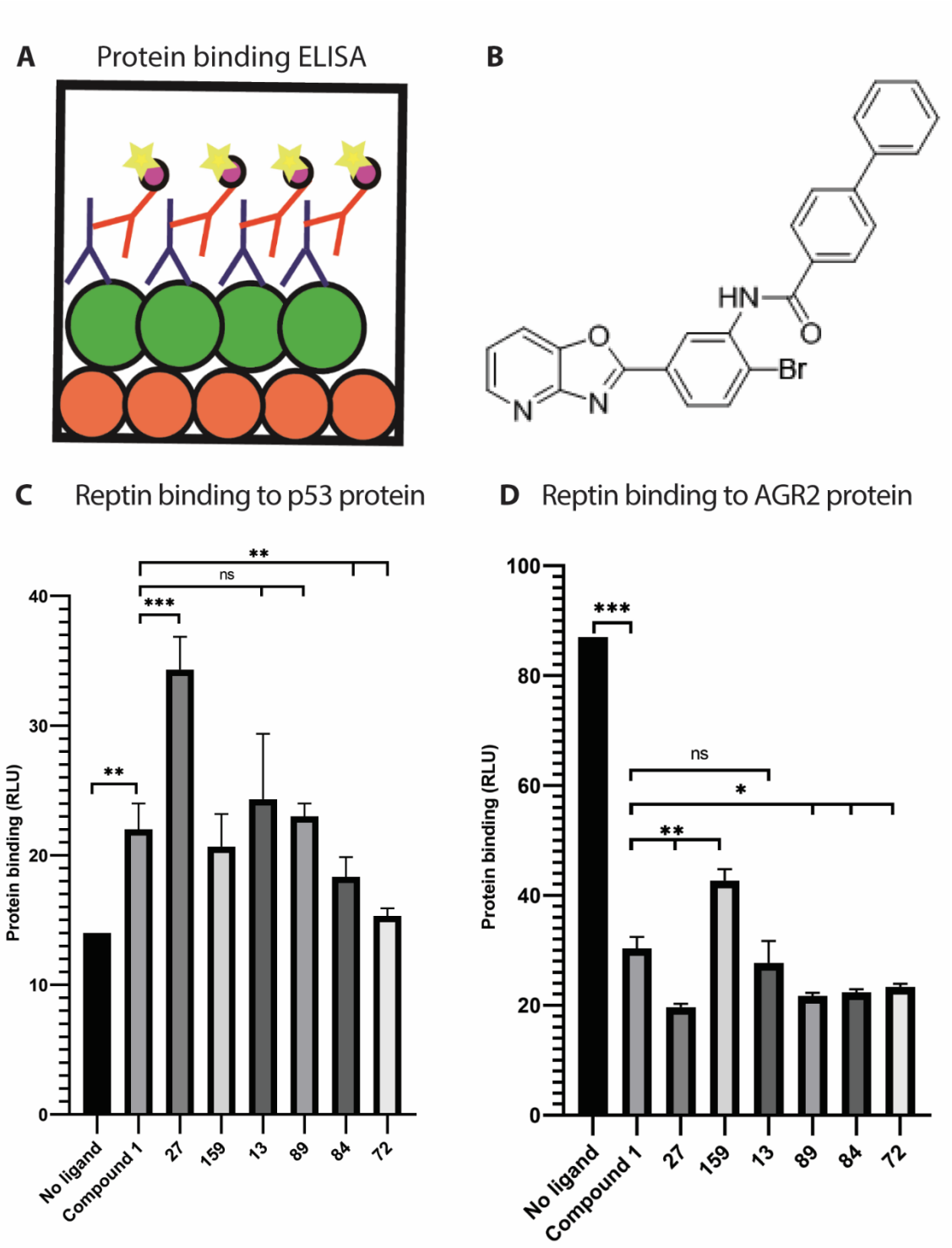


Figure 3.3 Protein-protein binding assays to measure changes in protein interactions
A) Schematic of protein-protein binding assay. The solid phase protein is shown in orange, the solution phase protein is shown in green. The primary antibody used against the solution phase protein is shown in blue and the secondary antibody conjugated to HRP is shown in orange with a pink HRP and a star representing the ECL reaction. B) Chemical structure of 27 (Liddean). C) Protein-protein binding assay showing changes in the ability of Reptin (solution phase) to bind p53 (solid phase) in the presence of a range of small molecules. D) Protein-protein binding assay showing changes in the ability of Reptin (solution phase) to bind AGR2 (solid phase) in the presence of a range of small molecules.

All the compounds used here are derivatives of Compound 1, with variations in the molecular structure of the molecule, as explained earlier (figure 3.1E). Compound 27 can be seen to have both the biggest increase in the amount of Reptin protein binding to p53 compared to the Reptin alone, and the compound which can be seen to have the greatest reduction in the amount of Reptin protein bound to AGR2 compared to Reptin alone. This was the first indication that this adaptation of Compound 1 could be of interest. When the results of these experiments were communicated with the collaborators it was clear that the composition of the “B” ring of Compound 1 was very important. It was variations here in the “B” ring which appear to have the biggest effect on the affinity of Reptin to binding its partners. From these experiments, it was clear that a single modification was the most effective. The structure of this compound can be seen 3.3B, this molecule different from Compound 1 as the halogen in the ‘B’ ring has been changed from chlorine to bromine. This small molecule from now on will be referred to as Liddean (Healy et al., 2015).

3.3.2.1.2 Peptide-protein assays

Peptide-protein assays are like protein-protein assays but here we are utilising overlapping peptide libraries to look at changes in the interactions between Reptin and short linear motifs. A schematic of this assay can be found in figure 3.4A. In brief, a polystyrene plate is coated overnight with streptavidin to allow binding of commercially synthesised biotinylated peptides (Biotin-SGSG-XXX…) to the well. Reptin was added to the wells and the protein was detected using a rabbit polyclonal antibody against Reptin, and a secondary antibody, swine anti-rabbit,

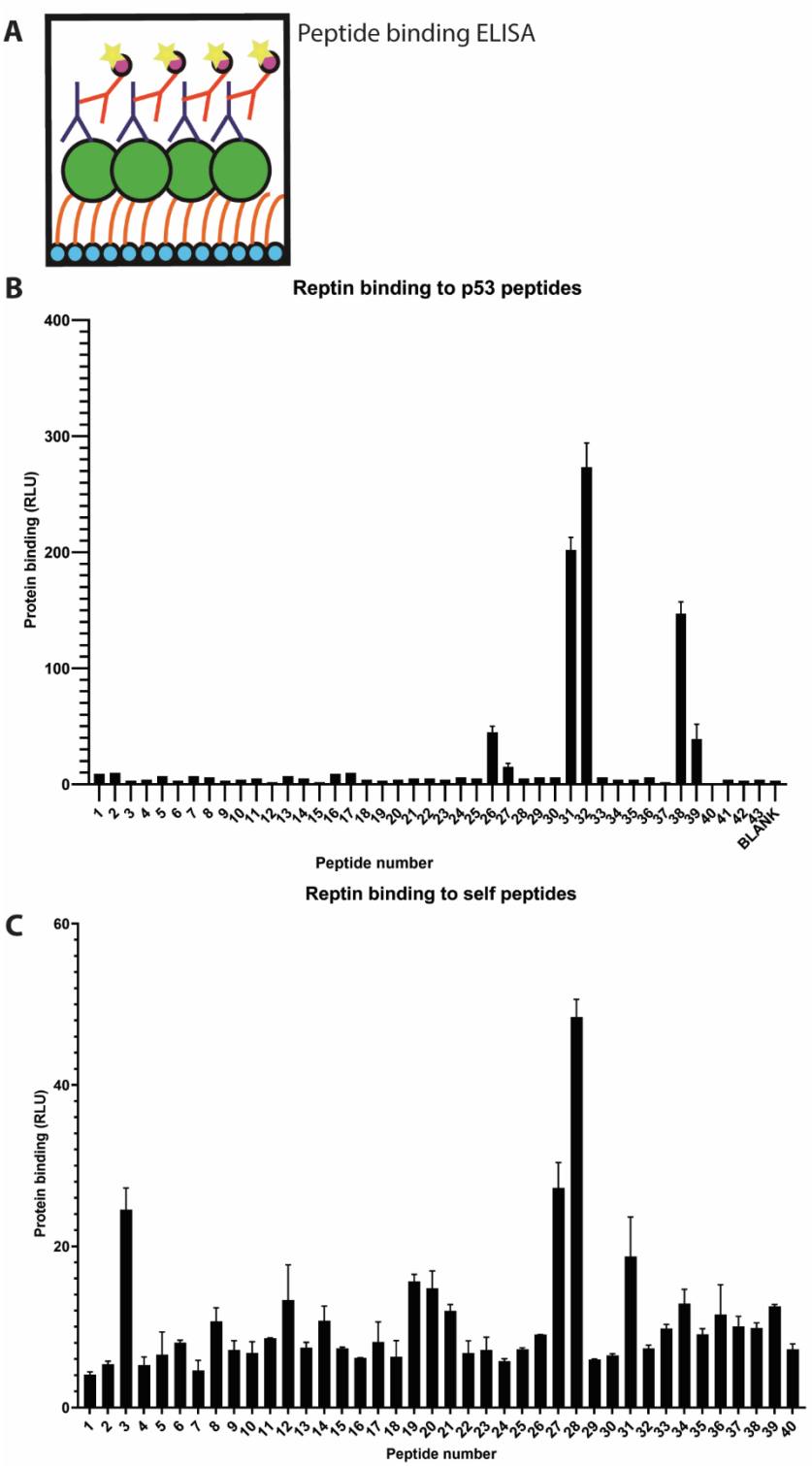


Figure 3.4 Binding of Reptin protein to peptides from p53 and itself
A) Schematic of peptide-protein binding assay. The streptavidin coated plate is shown in blue with the biotinylated peptide in orange. The solution phase protein is shown in green. The primary antibody used against the solution phase protein is shown in blue and the secondary antibody conjugated to HRP is shown in orange with a pink HRP and a star representing the ECL reaction. B) Peptide-protein binding assay using overlapping p53 peptides in the solid phase binding to solution phase Reptin protein. The blank well contains a mixture of all peptides showing background level binding of the Reptin antibody against the peptides. (Healy et al., 2015) C) Peptide-protein binding assay using overlapping Reptin peptides in the solid phase binding to solution phase Reptin protein. (Remnant et al., 2019)

using the ECL reaction measuring the luminescence in relative light units.

As the interaction between Reptin and p53 protein had previously been investigated in the lab (Healy et al., 2015), it was decided to look and see if there was specific peptide showing a high binding indicating a binding site for Reptin. The results of the experiment can be found in figure 3.4B. Here we can see two peptides showing a high binding to full-length Reptin protein, these are peptides 31 (SGSGRNSFEVRVCACPGRD) and 38 (SGSGYFTLQIRGRERFEMF). Peptide 31 can be mapped to the BOX-V domain of p53 and peptide 38 is mapped to its tetramerization domain.

As the only reliable commercially available antibody to Reptin is a rabbit polyclonal antibody, it is not possible to use this to investigate Reptin self-binding. To enable testing of this aspect of the behaviour of Reptin, a library of overlapping Reptin peptides was purchased which contain a 5 amino acid overlap at both N and C termini, creating a peptide of 15 amino acids in length. The initial experiment looked at how Reptin itself bound to these peptides and see which peptide or peptides had the greatest affinity of the Reptin peptide library. The largest signal is produced when Reptin is bound to peptides 3, 27 and 28. On inspection of the sequences of these peptides, it was decided that peptides 27-29 would be the most interesting to use for further experiments. This is due to their overlapping peptide sequences, which show very different binding strengths. The sequence of these peptides is as follows APVLIMATNRGITRI, GITRIRGTSYQSPHG and QSPHGIPIDLLDRLL (p27-29 respectively, also seen in figure 3.5A).

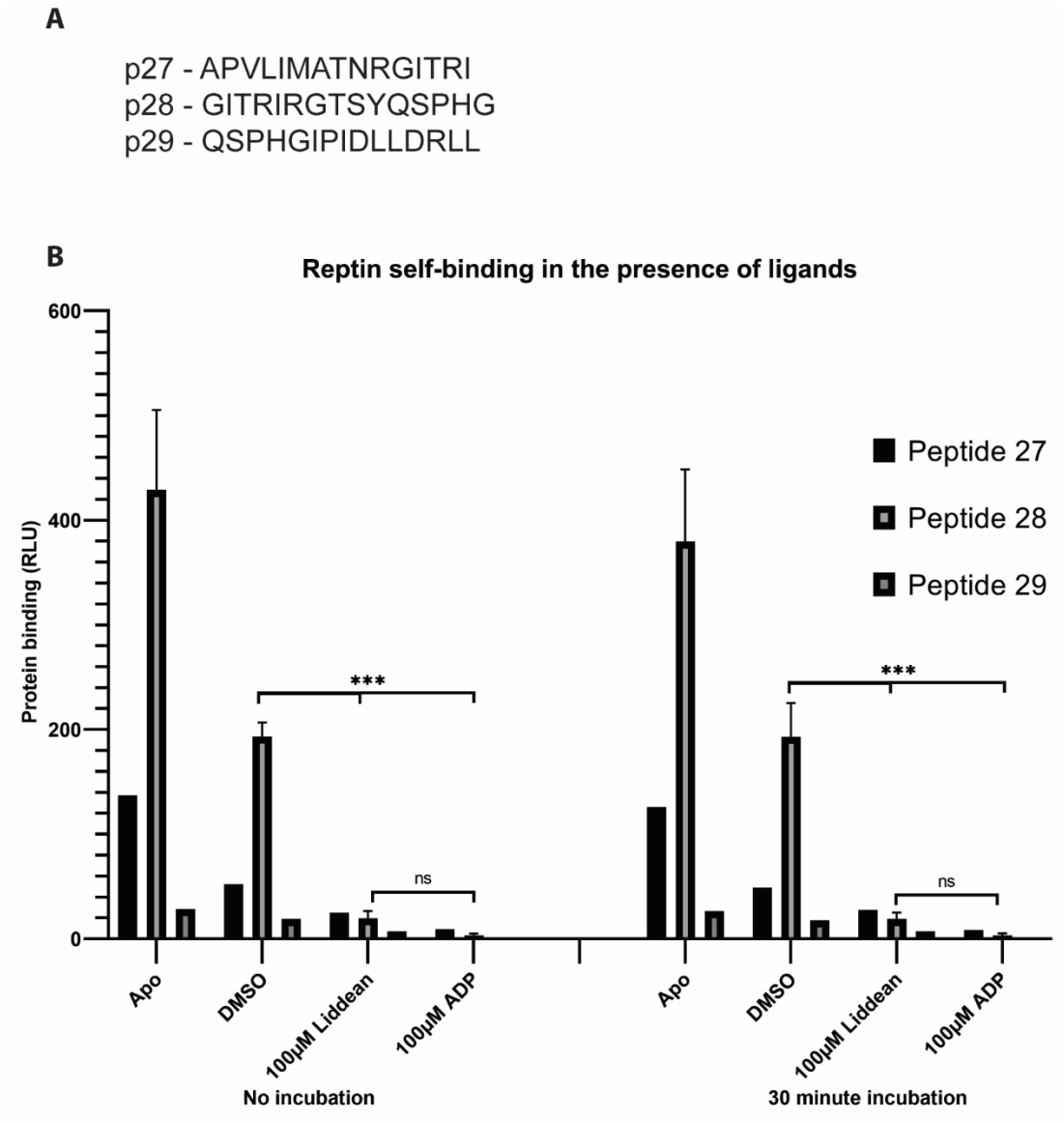


Figure 3.5 Measuring changes in self peptide binding on the addition of ligands
A) specific peptide sequences of Reptin self-peptides p27, p28 and p29. B) Peptide-protein binding assay using specific overlapping Reptin peptides in the solid phase binding to solution phase Reptin protein. In the first four conditions Reptin and small molecules were added at the same time to the peptides on the well, whereas the second four colours show Reptin which has been with small molecules described for 30 minutes prior to mixing with peptide of interest. (Remnant et al., 2019)

molecules affects the binding of Reptin to these peptides. As all of the small molecules are dissolved in DMSO it is a requirement to test whether the addition of the DMSO affects the binding affinity, which it appears to. The changes observed in Reptin exposed to small molecules in addition to DMSO are much more significant than those seen with Reptin and DMSO compared to Reptin alone (figure 3.5B). The first 4 sections show Reptin which has been added to the well at the same time as being exposed to the ligand, comparatively, in the second 4 sections the Reptin was incubated with the small molecules at 25ºC for 30 minutes before adding to the well. This shows the time required for the complex of protein and ligand to form and to see if there is. A requirement for any conformational changes to occur. Both datasets show very similar binding, this means that incubation is not required for complex formation, the binding between Reptin and the small molecules must occur very quickly, inducing conformational or oligomeric changes in Reptin. This assay also shows that the novel synthesised ligand, Liddean, acts very similarly in this assay to ADP, the molecule on which it was based.

3.3.2.2 Testing SDS resistance of Reptin oligomers

From the previous experiment, we saw changes induced by the presence of Liddean appear to be similar to those induced by ADP. As we know from other AAA+ proteins this is likely to be a result of oligomerisation. Here we are testing whether Liddean can affect the oligomeric state of Reptin. To achieve this, Reptin, apo and in the presence of ligands, was subjected to incubation with increasing percentages of SDS before being run on an SDS-PAGE gel, proteins were visualised using silver staining.

In the SDS incubation assay, increasing amounts of SDS are used to attempt to collapse the Reptin oligomer back to a monomer, the more stable the oligomeric

structure is, the greater the percentage of SDS required to collapse it. SDS is a molecule known to reduce proteins to their monomeric state by coating the polypeptide in SDS allowing it to flow easier through the polyacrylamide gel. Here the gel is stained using the silver stain protocol to detect the very low amount of protein in all lanes of the gel. In figure 3.6A we can see that, in lanes where Liddean in present, but not ADP, there are large higher-order structures which can are stable in the presence of SDS, due to both their different sizes and discrete banding they may represent different confirmations of the hexametric Reptin. In all lanes a band is visible around 130kDa, this could be a dimeric form of Reptin.

3.3.2.3 Crosslinking of Reptin oligomers

Crosslinking of Reptin in the absence and presence of Liddean with increasing amounts of glutaraldehyde should give more detail about the higher-order oligomers visualised in the SDS experiment. In figure 3.6B much less monomeric Reptin is present in the samples containing Liddean even at the lowest concentration of glutaraldehyde. This likely means that Liddean binding to Reptin induces its oligomerisation and creates a more stable form of the oligomer than its ligand.

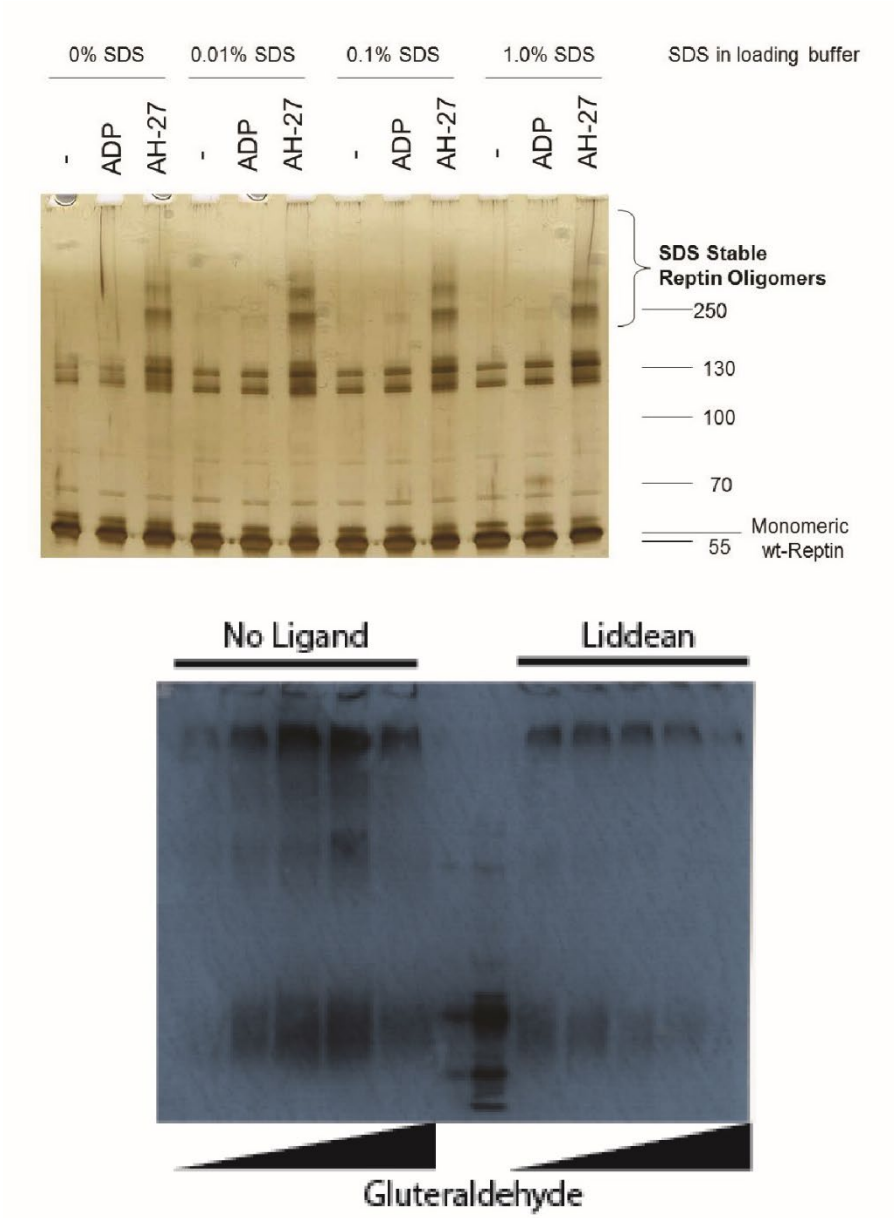


Figure 3.6 Examining changes in oligomeric state induced by small molecule binding

A) SDS PAGE silver stained gel showing Reptin incubated in presence of small molecules in increasing concentrations of SDS. (Healy et al., 2015) B) Western blot showing Reptin crosslinked with glutaraldehyde in the absence and presence of Reptin

3.3.3 **Testing Compound 1 analogues in cells**

To measure changes in cells caused by the addition of ligand, these must be introduced into cells in a reproducible manner. Due to the insolubility of Liddean in aqueous solution, it was required that Liddean must remain in dissolved in DMSO, as with the previous experiments. Previous work has shown that it is possible to introduce small molecules into cells using transfection reagents which would normally be used for DNA and RNA (Rena et al., 2004).

3.3.3.1 Measuring changes in cell impedance on the addition of ligand

The use of non-invasive electrical impedance monitoring systems allows for the quantitative observation of cell proliferation and changes in cell morphology with high temporal control, in real-time. Cell impedance is as a measure of cell adherence and proliferation. The RTCA iCELLigence is designed to do just that. Cells are grown in wells on top of electrodes, a small electrical current is passed across and changes in this current are measured and plotted against time.

Here an RTCA iCELLigence instrument was used to measure changes in cells caused by the addition of Liddean. For these experiments, the HCT116 cell lines were chosen. These are a human colon cancer cell line which is available as p53 positive and negative. Due to the known binding of Reptin, AGR2 and p53, these interactions and the changes seen as a result of the presence of Liddean in vitro, it was thought these cell lines might show differences in impedance due to the presence of active of p53 in the cell. The cells were incubated in the chambers for 24 hours, allowing for normalisation between experiments and checking the number of cells in each well same in all. After 24 hours, Liddean was added to cells at different concentrations (or DMSO alone) after incubation with Attractene as a carrier for the small molecules. The cells were monitored for a further 72 hours.

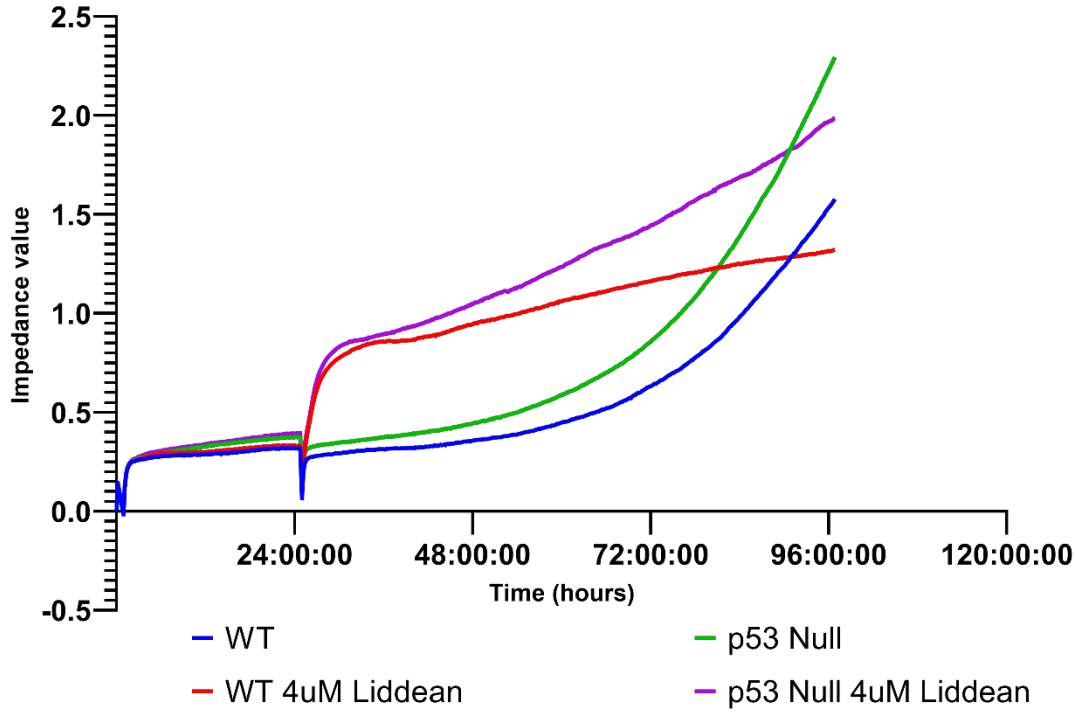


Figure 3.7 iCELLigence analysis of HCT116 cells exposed to Liddean
Time course of cell culture impedance values. Cells were incubated for 24 hours in wells before addition of 4uM Liddean.

once exposed 4 μM Liddean. The dose of 4 μM Liddean shows a dramatic increase in impedance on the addition of the small molecule. Due to the speed of this increase, it is unlikely that this is due to cell division as the cells are not synchronised and therefore not ready to enter mitosis. This increase is likely due to cells adhering to the surface of the well and changing their morphology after exposure to Liddean. At this stage, it was not possible to confirm this hypothesis and further testing will be required.

3.3.3.2 Testing cell cycle changes due to the addition of ligand

Fluorescence-activated cell sorting (FACS) is a specialized type of flow cytometry which counts the number of cells in a population that contain a particular amount of DNA, thus enabling identification of the percentage of the cell population is at particular stages in the cell cycle or apoptosis. FACS can be used to ascertain whether the observed increase in cell impedance was due to the addition of Liddean causing cells to synchronously enter mitosis. The test was carried out 6 and 48 hours after the addition of different concentrations of Liddean. The 6-hour time point was chosen as in the iCELLigence experiment this is the time after addition of Liddean where I saw the greatest increase in cell impedance value. The 48-hour time points should allow the cells to go through two cell cycles.

The results of these experiments can be found in figure 3.8. Here, we can see there is no observable difference between treated and untreated cells at any of the timepoints observed, all data here is not significant. The total number of cells in each gate is seen to be similar, independent of the condition they come from, this shows that it is also unlikely that the addition of Liddean is inducing an increase in apoptosis in these cells.

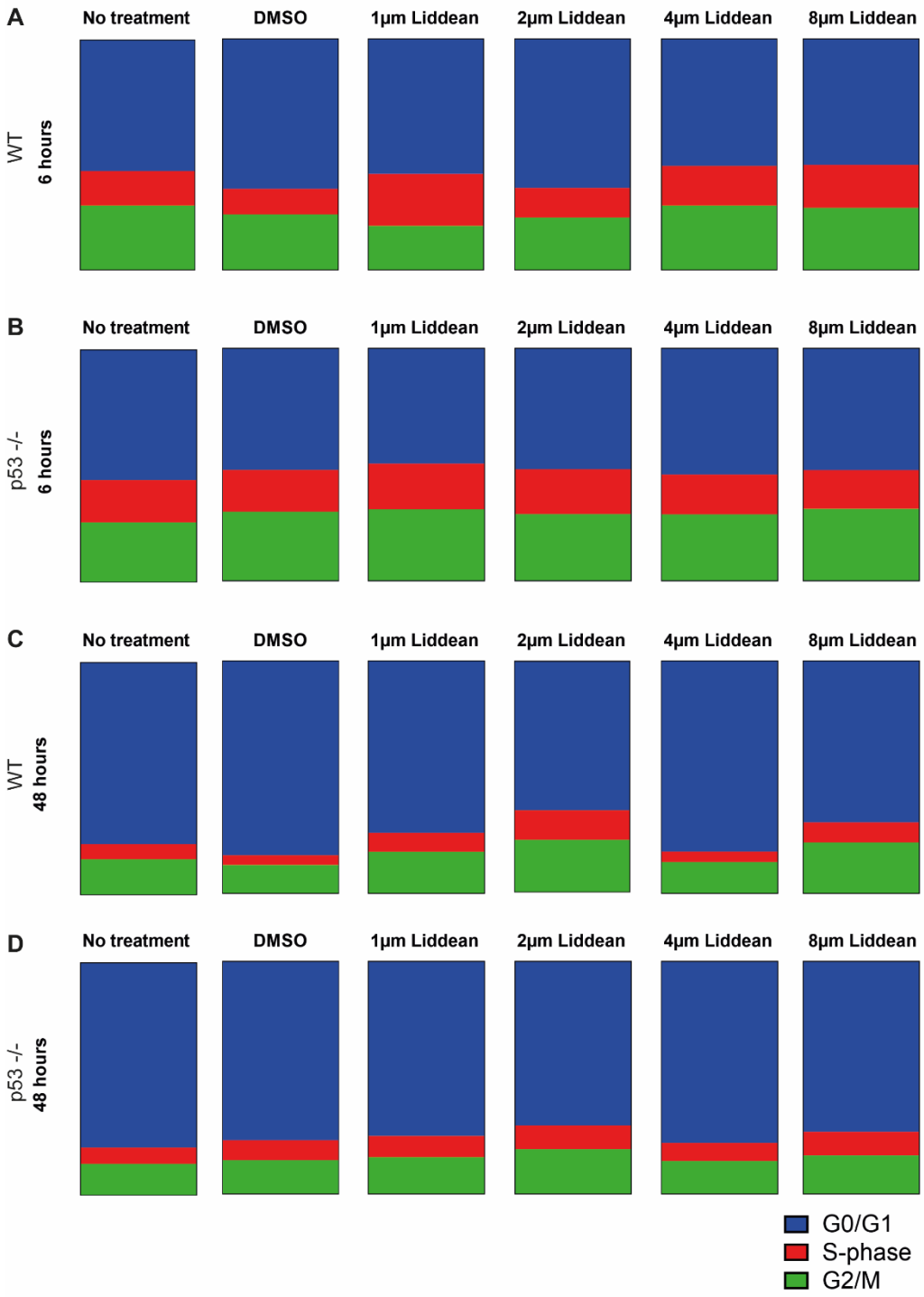


Figure 3.8 FACS analysis of cells exposed to Liddean
6 and 48-hour time points looking specifically at cell cycle stage in HCT116 WT and p53 -/- cell lines.

3.4 Discussion

The AAA+ protein family are characterised by their nucleotide-binding that induced formation of oligomeric complexes. This investigation exploited this ability of Reptin to develop a small molecule binder of Reptin which can be used to explore the conformational and oligomeric changes in Reptin further. As ATP is highly abundant and required for a vast number of cellular functions it is very difficult to examine the changes of a specific protein-ATP/ADP interaction in vivo. The specificity of this new small molecule might mean it is unable to bind into the ATP binding sites of some other AAA+ proteins due to steric hindrance from amino acids outside of the pocket, although further investigation is required.

The ability to add a specific amount of ligand which is not able to undergo hydrolysis enables us to study the ligand-bound form of Reptin without the conformational changes which occur as a direct result of the ATPase activity of the protein. The information gained with the use of the mimetic can be compared to that found when using either ADP or ATP. This will allow for further dissection of the complex nature of this protein.

The initial protein-protein interaction studies show how ligand binding, and therefore likely conformation or oligomeric change in the protein of interest, have a direct, and specific influence on the binding affinity of the protein of interest. The experiments show that the binding of p53 and AGR2 must occur via different mechanisms as the binding of one is increased with ligand addition whilst the other is decreased. The same trend is seen with some of the other ATP mimetic small molecules. These observed changes were similar to those from the peptide-protein Reptin peptide library assay. Here we observe an increase in binding of specific self-peptides,

including a specific small linear motif, which is directly affected by the addition of small molecules.

The studies looking for higher-order oligomers of Reptin in the presence of ligands both show the existence of a stable high molecular weight protein. The SDS assay shows that it is possible the binding of Liddean to Reptin is more stable than the binding of Reptin to ADP as the high molecular weight proteins are only SDS resistant in this condition. The crosslinking with glutaraldehyde shows that wildtype Reptin can be crosslinked into higher molecular weight oligomers in the absence of ligand, however, the presence of ligand appears to reduce the amount of monomer visualised to a greater extent showing similarly that the higher-order oligomer formed with the addition of Liddean is stable.

As we were able to discount changes in impedance due to the cell cycle with the FACS data, we can conclude that that increase in impedance seen in the iCELLigence experiment is due to a change in cellular adherence or morphology with the addition of Liddean. In this instance, this may show the off-target effects of Liddean. As Liddean is an ATP mimetic designed to bind to a Walker A motif other proteins within the cell are likely able to bind it if they contain this motif. Liddean could bind to any of the P-loop NTPases which also contain Walker A sites, many motor proteins are found in this superfamily (Berg et al., 2002). The members of the AAA+ superfamily also fall into the P-loop NTPases superfamily. P-loop NTPases include myosin and kinesins (Leipe et al., 2002). Dynein is a likely candidate as an off-target binder as it is a member of the AAA+ superfamily and has roles in cytoskeletal regulation (Kim et al., 2012) although there are several others also (Vale, 2000, Hanson and Whiteheart, 2005, Snider et al., 2008). As bioelectrical impedance cell-based biosensors are so sensitive, they can be used to screen bioactive agents. They easily translate changed in cell growth into a measurable

figure and can be used to determine the efficacy of a specific compound (Ramasamy et al., 2014). There is also evidence that binding of small molecules to protein targets *in vitro* can cause changes in the measured electrical impedance of a solution, this could mean that the change in impedance measured is a direct result of Reptin and Liddean binding to one another although this would need further investigation (Pisa et al., 2019). Although the reason for the change in impedance value in still unknown in this case, this result still gives us additional information into the action of Liddean within the cell. The results in the iCELLigence experiment are like what would be expected during attachment of cells on to the plate so is it possible that Liddean is affecting the adherence of the cells.

The use of Liddean may enable discovery of the complete protein-protein interaction network of Reptin. Small molecule stabilisation or destabilisation of specific interactions is underused used in the clinic, especially those which act on small linear motifs in protein interactions. Currently drug leads with similar are emerging for use in the clinic, one if which targets the linear peptide motif binding groove of MDM2 oncoprotein (Ray-Coquard et al., 2012). Linear peptide motifs are interesting due to their combination of specificity and lack of study. This aspect of protein-protein interaction could be used in new pathways of drug discovery as currently, this dynamic system is yet to explore.

Several other ATP mimetic molecules have been developed with the hope that they are highly specific for their target. As ATP is so ubiquitous in the cell, it is complex to ensure that the small molecules designed to bind into the binding sites on these proteins are specific and do not have several off-target effects. Many kinase inhibitors have been developed as ATP mimetics, in the past, these were designed and then discovered to target multiple enzymes, although this was through to be positive. However, more recently an increase in specificity has been required as the

understanding of protein interaction pathways is revealed (Knight and Shokat, 2005). There have been many ATP competitive small molecules developed as phosphatidylinositol-3-kinase (PI3K) inhibitors, this is due to PI3K having a prominent role in many cancers due to the Akt/mTOR pathway. Currently, eight PI3Ks that have been identified and these small molecules have varying binding to each of the proteins (Liu et al., 2017). This is an important example in drug development as the 'off-target' binding to proteins which are highly similar to the protein of interest can cause unwanted effects. The methodology used here, with the use of *in silico* screening, enables comparison of binding between a single small molecule and many binding pockets could help in the future for more specific drug targeting. Recently, an ATP mimetic called Sorafenib has been developed as an inhibitor of Reptin (Nano et al., 2020), however, this small molecule does not appear to have an affect on the overall conformation of Reptin as we see with Liddean. This inhibitor is known to reduce ATPase activity. It would be interesting to compare the affects of the two small molecules.

Other AAA proteins have also had small molecules developed with Walker A and B binding in mind. Using the previously published crystal structures, in a similar way to which LIDAEUS does, small molecules have been developed which should have able to avoid resistance by looking at possible mutations which could occur in the binding pocket of proteins, specifically spastin (Mitchell and Matsumoto, 2011).

The ability to target small molecules to specific protein-protein interactions has been further advanced by the use of tools like LIDAEUS. As we are now able to use virtual screening to test whether a particular small molecule is likely to bind to a specific interface, it is possible to look at the interfaces between two proteins in the same complex which would normally be thought of as being inaccessible and design small molecules which will bind directly, blocking interaction (Wu et al., 2019). This

use of super-computing as part of drug development is increasing in popularity due to its speed and reduction in costs (Mitchell and Matsumoto, 2011). Focusing on a specific protein-protein interaction gives a greater standing of the likely side effects and possible efficacy of blocking a particular stage in the pathway (Fotis et al., 2018). These larger screening mythologies also allow for drugs which have already been developed and have gained approval to be tested in silico for use with other interactions and diseases.

Chapter 4 Developing antibodies to distinguish between different oligomeric states of Reptin

Some of the work in this chapter was published in 2015 in Chemical Science as a collaboration between the laboratories of Ted Hupp and Nicholas Westwood (Healy et al., 2015).

4.1 Introduction

Antibodies are often used to identify proteins in vivo and in vitro, in both research and medicine. There are many biochemical assays which use antibodies to identify the presence of a protein of interest, such as in ELISAs assays (Engvall, 1980), or localisation of a protein of interest, such as immunofluorescence (Donaldson, 2015). Antibodies are also used in a clinical setting for both diagnostic and treatment purposes. For diagnostic purposes, antibodies can be used to detect particular proteins within bodily fluids, for example, human chorionic gonadotropin is usually present in the urine of pregnant women during the first trimester (Butler et al., 2001). Another use of antibodies in a diagnostic setting can be to detect a patients' immunoglobulins which can be indicators of specific illnesses and diseases. The therapeutic use of antibodies is also widespread, they can be used to treat several diseases including arthritis, chon's disease and cancers (Scott et al., 2012) (figure 4.1).

Development of many of these biological techniques and therapeutic uses is made is easier when there are multiple antibodies for a single protein of interest. Monoclonal antibodies have defined single epitopes on an antigen, therefore these can be used to look at changes in conformation or protein interaction if their binding site is masked in these different situations. Polyclonal antibodies bind to several epitopes on a single antigen, this can be very helpful in protein interaction studies if

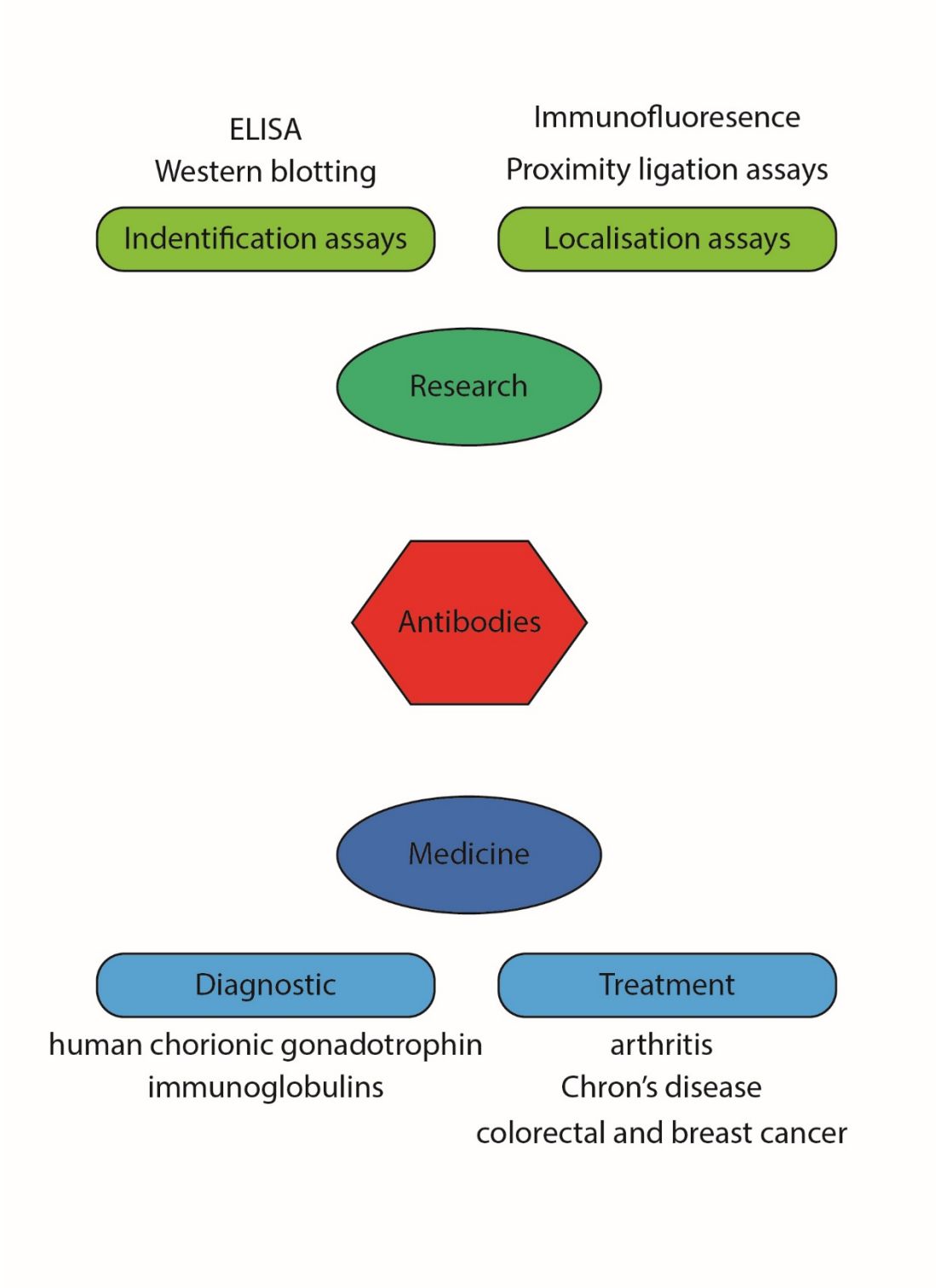


Figure 4.1 Different uses of antibodies.
Green identifies the research uses. Blue identifies the medical/pharmaceutical uses.

bind due to the vast number of epitopes, allowing for constant detection of the protein of interest. These are visualised in figure 4.2A.

4.1.1 The molecular structure of antibodies and their derivatives

Antibodies are large proteins, around 150 kDa in total. They are composed of 4 polypeptide chains, two heavy and two light chains, each containing variable and constant regions, creating the 'Y' shaped molecule (figure 4.2B). More recently, the development of alternatives to antibodies has been carried out, these include the production of Fab fragments, single-chain variable-fragments (ScFv) and VH/VHH (figure 4.2C). Fab fragments are the largest of these derivatives and contain the variable regions as well as constant regions as a linker. ScFv is smaller, normally between 25 kDa and 35 kDa, and contain just the variable regions. VH/VHH contain just a single variable region (Bird and Webb Walker, 1991). These smaller derivatives have several advantages over full-sized antibodies including being able to be synthesised in bacterial rather than animal systems. Another advantage of ScFv and VH/VHH is that they lack the constant regions which tend to be specific to the animal in which the antibodies are made in, and therefore are more likely to be identified as foreign by the immune system.

ScFv can be used in a form of phage display where the variable fragments are expressed on the protein coat of bacteriophage. In phage display, a library of these bacteriophages, each expressing different ScFv on its protein coat, can be used to determine the number of ScFv that have an affinity for a protein of interest. Once the selection has occurred, the ScFv can be sequenced and then developed for use in other protocols, such as immunofluorescence, ELISA and PLA (Jain et al., 2016).

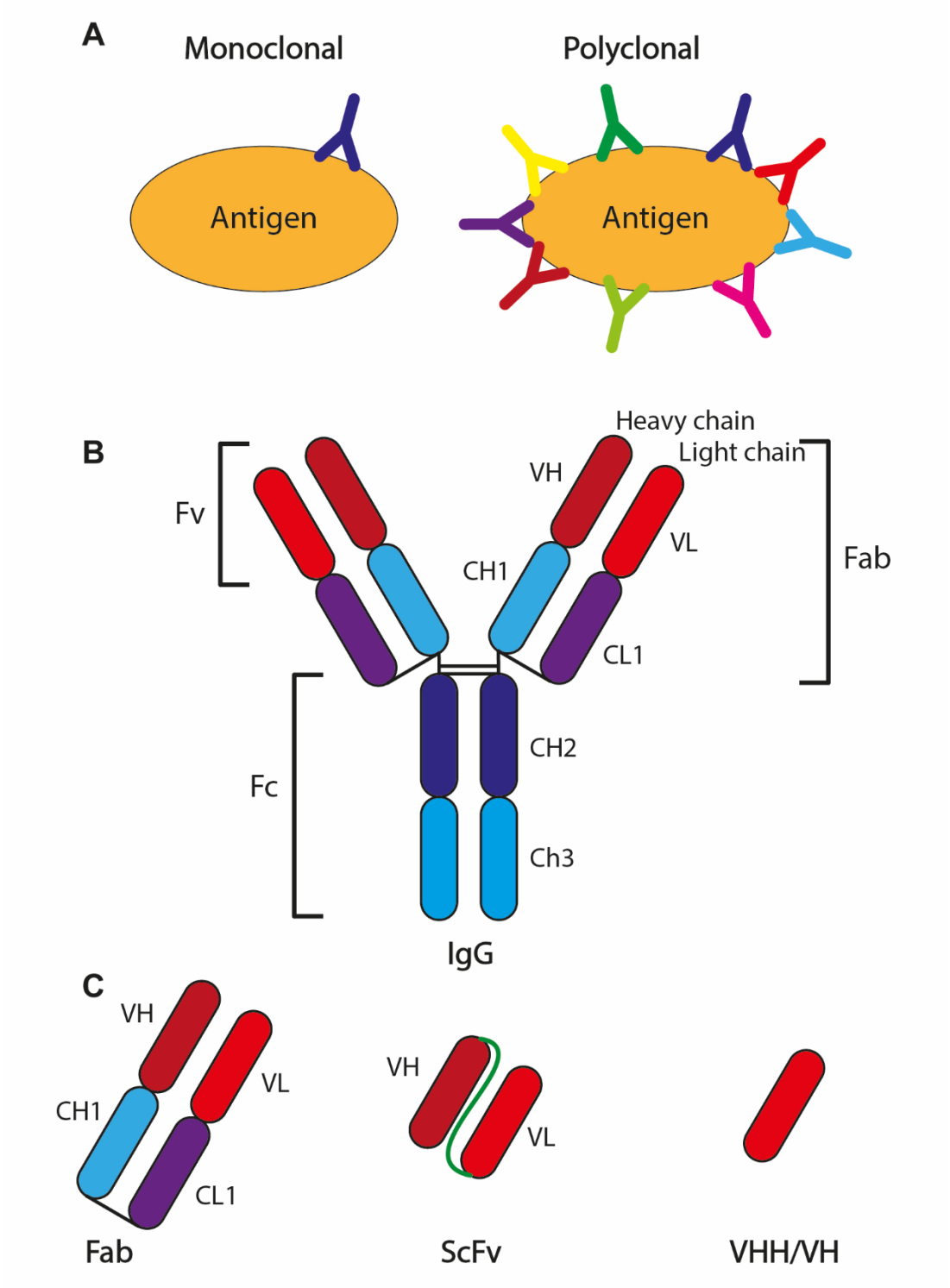


Figure 4.2 Schematics of antibodies
A) Difference between a monoclonal and polyclonal antibody. B) Schematic of full-length antibody containing a heavy and light chain. C) Difference between Fab fragments, ScFv and VHH/VH

The first monoclonal antibody therapy was licenced in 1986 (Liu, 2014),m since then many more have come to the market, currently, there are over 80 (Bates and Power, 2019) in use, many of which have roles in cancer for diagnostics and therapeutic purposes (Scott et al., 2012). However, there are several commonly occurring problems with these including clearance times from the bloodstream, non-specific binding to normal organs and tissues and low efficacy. Due to their much smaller size, ScFv exhibits unique properties which mean they are not subjected to the same issues which affect larger antibody fragments and whole antibodies. ScFv is much less likely to be recognised and removed by the immune system (Kholodenko et al., 2017). Also, they are much easier to produce due to their ability to be expressed and purified from bacteria in high volumes at lower costs (Fernandes, 2018).

4.2 Aims

The experiments in this chapter aim to develop tools to help measure a change in oligomeric state. In Chapter 3, Liddean was developed and it was shown that this small molecule was able to induce an oligomeric change in the Reptin protein. We know that Reptin and Pontin form complexes in cells and this is well documented, but it is currently not possible to visualise oligomeric Reptin.

We want to investigate:

- If it is possible to create antibodies able to distinguish between the oligomeric forms of Reptin
- If we can develop single chain antibodies which are specific to Reptin when bound to specific partners, either small molecules or other proteins
- If we can use these single-chain antibodies in current methodologies, such as proximity ligation assays, to determine the oligomeric state of Reptin inside cells and whether this is localisation specific

4.3 Visualising Reptin and its binding partners in cells

In the previous chapter, we looked in detail into the interactions of Reptin and its known binding partners in vitro. Using a combination of immunofluorescence and proximity ligation assays we can observe the same interactions in cells. Visualization of Reptin and its known binding partners has previously been studied in the lab (Healy et al., 2015).

4.3.1 Localising protein interactions with Proximity ligation assays

To confirm the interaction between Reptin and its binding partners' proximity ligations assays were utilised. In this experiment, a signal is only observed when the two proteins of interest are in very close proximity with each other either directly interacting or as part of the same complex (Clausson et al., 2011, Gullberg et al., 2004). The interactions between Reptin and Pontin and Reptin and p53 can all be confirmed in vivo using this technique.

In brief, a cell was grown on coverslips and then transfected with either DMSO or Liddean with the Attractene carrier (Rena et al., 2004) (as seen in Chapter 3, figure 3.7) and incubated for 24 hours. Cells were fixed with 4% paraformaldehyde. The samples are permeabilised before unspecific antibody binding is blocked with the addition of 3% BSA in PBS. Primary antibodies raised against your target antigens are incubated with the samples, one polyclonal and one monoclonal raised in different animals. PLA probes were added to cells and incubated to 1 hour at 37°C. The ligation mix was added for 30 minutes at 37°C followed by the polymerase amplification mix for 2 hours at 37°C. Coverslips are then mounted onto slides before imaging using a BX51 (Olympus) Fluorescent microscope.

We can visualise foci which are a result of the interaction between Reptin and p53 (figure 4.3B, control seen in 4.3A), here we can see that the foci are almost exclusively present in the nucleus of the cell, this correlates with previously published data on the localisation of these two proteins. We could not visualise Reptin and AGR2 at the same time due to the antibodies available, for three-channel imaging, the primary antibodies must be raised in different animals.

When investigating the effect the presence of Liddean has on the protein-protein interactions of Liddean we looked specifically at the Reptin-Pontin interaction. Here we can see that in the absence of Liddean, Reptin-Pontin foci are found in the nucleus of the cell but after 24 hours exposure to 2mM Liddean results in the shuttling of these foci into the cytoplasm of the cell (figure 4.3C). The quantification of this experiment is seen in figure 4.3D (grey bars – cytoplasmic, white bars - nuclear), here we can also see that 2mM Liddean appears to reduce the total number of Reptin-Pontin foci but in 1mM Liddean they remain at a similar level to the control; however, we still see the shift of the foci into the cytoplasm of the cell. The number of foci was counted by hand, using the DAPI staining to qualify whether a spot was nuclear or cytoplasmic.

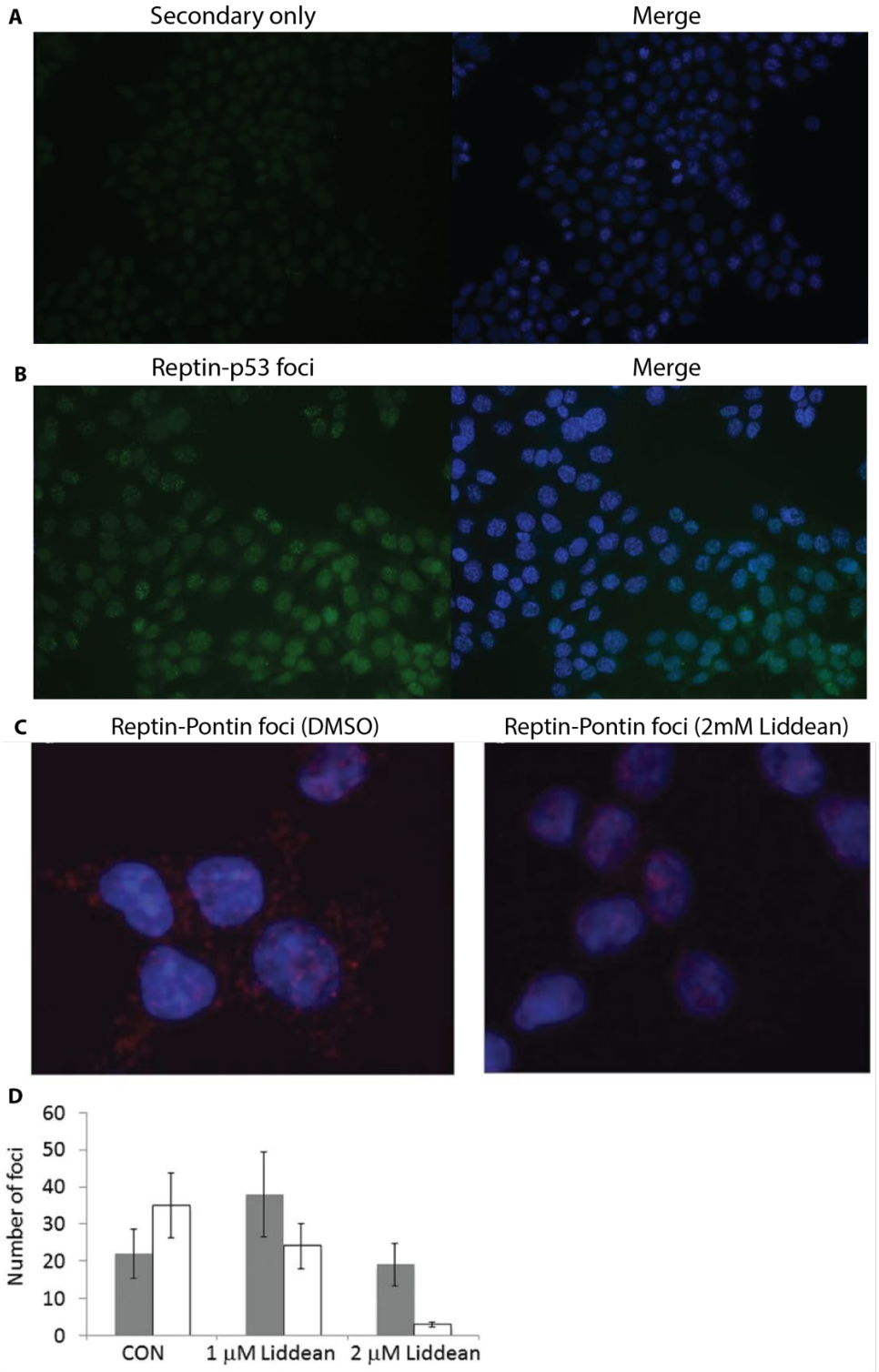


Figure 4.3 Proximity ligation assays (PLA) between Reptin and various binding partners in HCT116 cell line.

A) Control PLA using no primary antibodies. B) PLA of Reptin-p53 interaction, PLA foci in green channel with DAPI used to localise nuclei. C) The effect of Liddean on Reptin-Pontin interaction. DAPI used to localise foci as nuclear or cytoplasmic in either DMSO or 2mM Liddean treated cells. D) Quantification of number of Reptin foci localised as either cytoplasmic (grey bar) or nuclear (white bar). (Healy et al., 2015)

4.4 Antibody phage display to produce single chain antibodies for Reptin analysis

Development of a single-chain variable-fragment (ScFv) library specific to Reptin will allow for more robust investigation of Reptin in vivo and in vitro. Due to a combination of epitopes, it should be possible to determine different oligomeric states of Reptin.

4.4.1 Development of Reptin specific ScFv

A single-chain variable-fragment (ScFv) library expressed on the coat of bacteriophage was created previously in the lab (Jain et al., 2016). A novel library of ScFv was extracted from naive canine spleens using extracted RNA. PCR was used to amplify the heavy and light variable regions before being ligated into phage display vectors. The Phagemid vectors were then transformed into *E. coli* for the production of the bacteriophage expressing the ScFv on its protein coat. The resulting ScFv bacteriophage library was used to aid the creation of a Reptin specific ScFv library.

A schematic of the panning process can be seen in figure 4.4. In brief, Reptin and AGR2 were bound to wells overnight. The library was precleared against the opposite condition, to remove ScFv which had a high binding affinity to protein in general, this was incubated for 1 hour. Here I did not use Liddean as I wanted to create a library of ScFv able to recognise all oligomeric forms of Reptin, not just the form induced by Liddean binding. The precleared library of phage was added to the well containing the protein of interest and incubated for 1 hour. Unbound phage was washed away and then the bound phage is eluted. The eluted phage was amplified and then used for a second round of panning in the place of the library. A third-round was also carried out, in the same manner, using the amplified phage from the second round as the input.

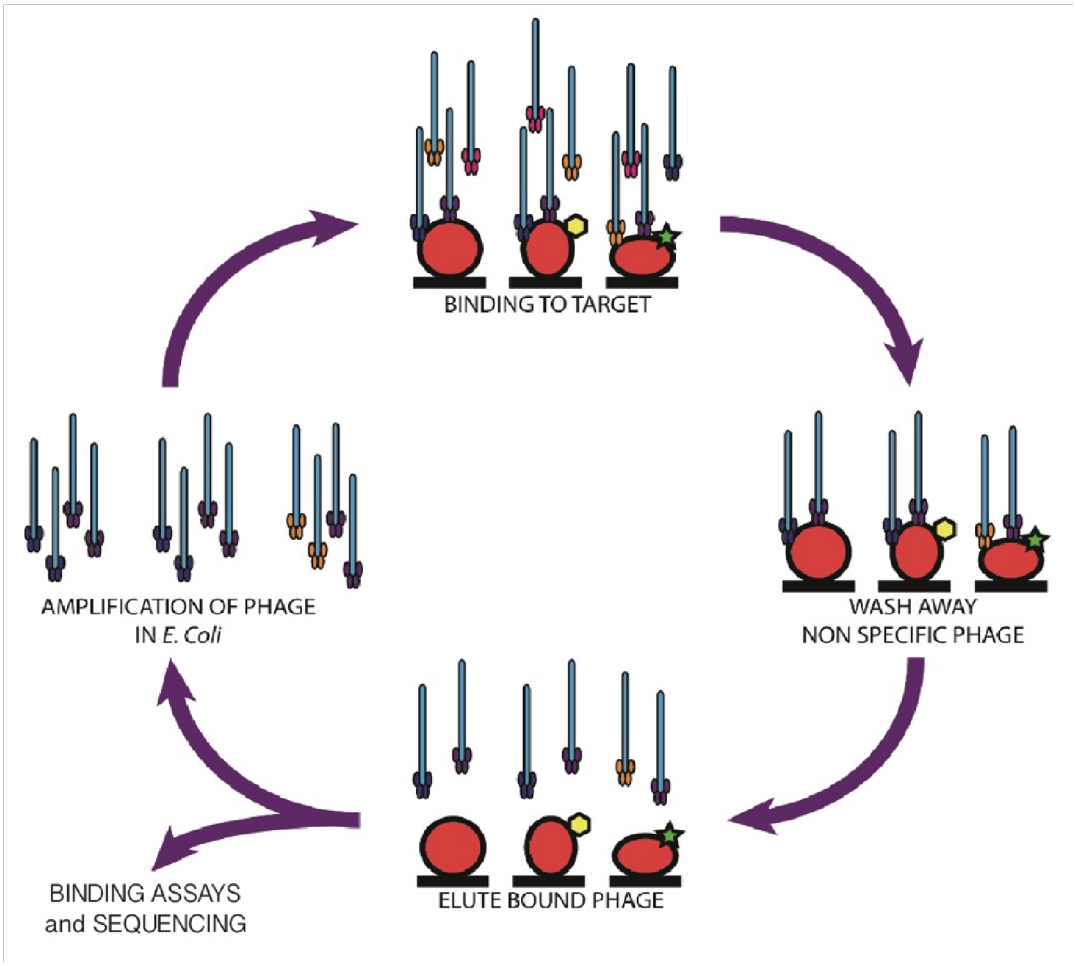


Figure 4.4 Panning process for antibody phage display.
Cycle of panning showing rounds of amplification and elution used to develop ScFv. Different expressing ScFv on the protein coat can be seen as different coloured circles. The protein of interest is seen as a red ball, bound to different ligands (yellow hexagon and green star) resulting in different conformations of the protein of interest.

Once three rounds had been carried out, the affinity of the bacteriophage for Reptin was carried out using AGR2 as a negative control to check the specificity of the phage pools (figure 4.5). Here we can see the specificity of the pools increased over the rounds and that there is a reduction in the amount of phage able to bind Reptin in the presence of Liddean compared with Reptin only. This was encouraging and could mean that there is conformation-specific ScFv in the polyclonal pools.

To develop ’monoclonal’ ScFv, the round 3 polyclonal pool for Reptin was used to infect bacteria before being plated out to hopefully form colonies expressing a single ScFv. 96 colonies were picked and grown-up similarly to the polyclonal pools before being tested for binding against Reptin and Reptin + Liddean. The results of these experiments can be seen in figure 4.6 showing the three graphs containing the binding of all 96 ‘monoclonal ScFv’, 1-32, 33-64 and 65-96.

The ‘12 top binders’ were then tested in a protein-protein binding experiment using the ScFv to detect Reptin bound to AGR2 (figure 4.7). This allowed observation of whether these ScFv could be used in established protocols. Many of the ScFv was seen to be able to detect the presence of Reptin in this protein-protein binding assay.

Previously designed primers were to extract the ScFv sequences for sequencing by Source Bioscience. From the 12 available ScFv, 6 produced PCR products that gave readable sequences, these can be seen in figure 4.8. From these 6 sequences, we can see that we had purified 5 different ScFv, 21 and 59 have identical sequences in both the heavy and light chains but all others have different sequences. The FR regions are framework regions that do not have variation, but the CDR regions are the variable regions which are responsible for binding. There is also a linker which combines the heavy and light chain variable fragments, this is

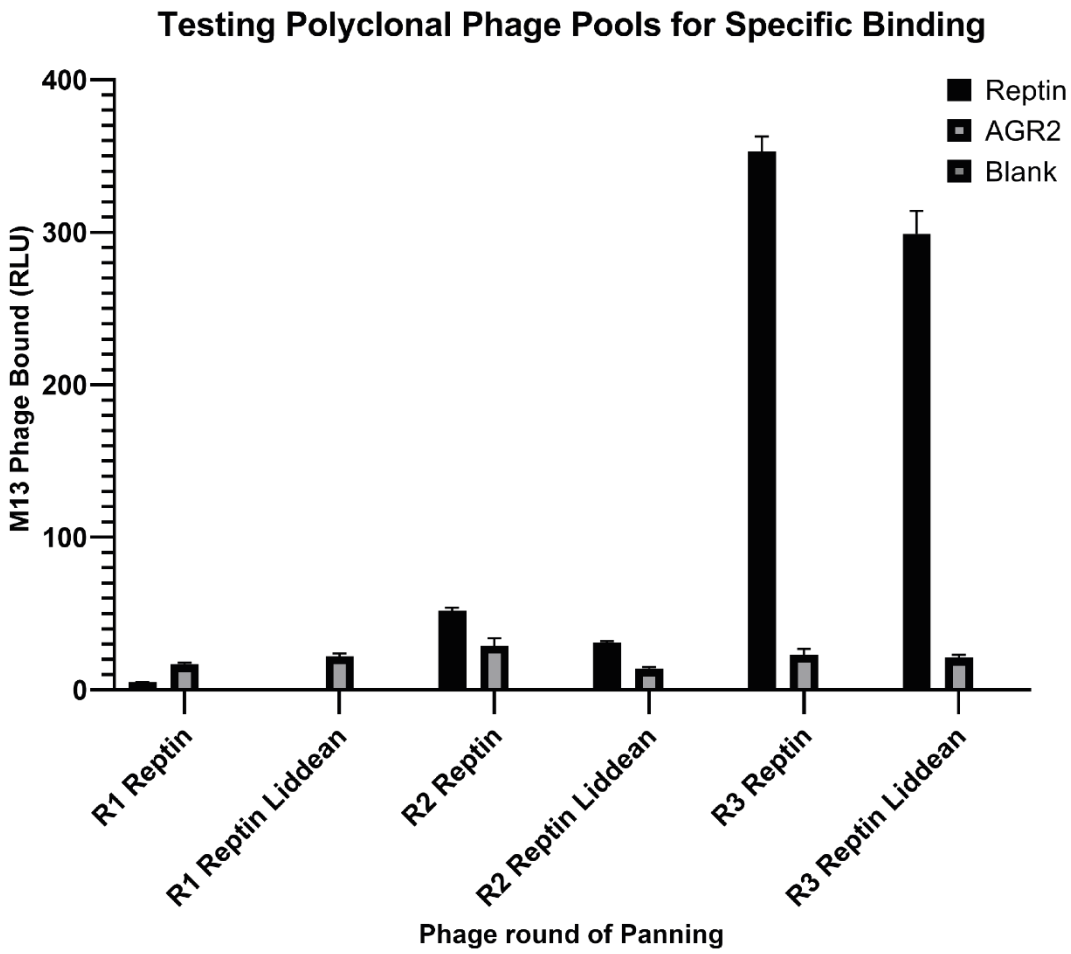


Figure 4.5 Testing Polyclonal Phage Pools for Specific Binding
Testing ScFv Polyclonal phage pools from multiple rounds against Reptin and AGR2 (used as control protein) in presence and absence of Liddean.

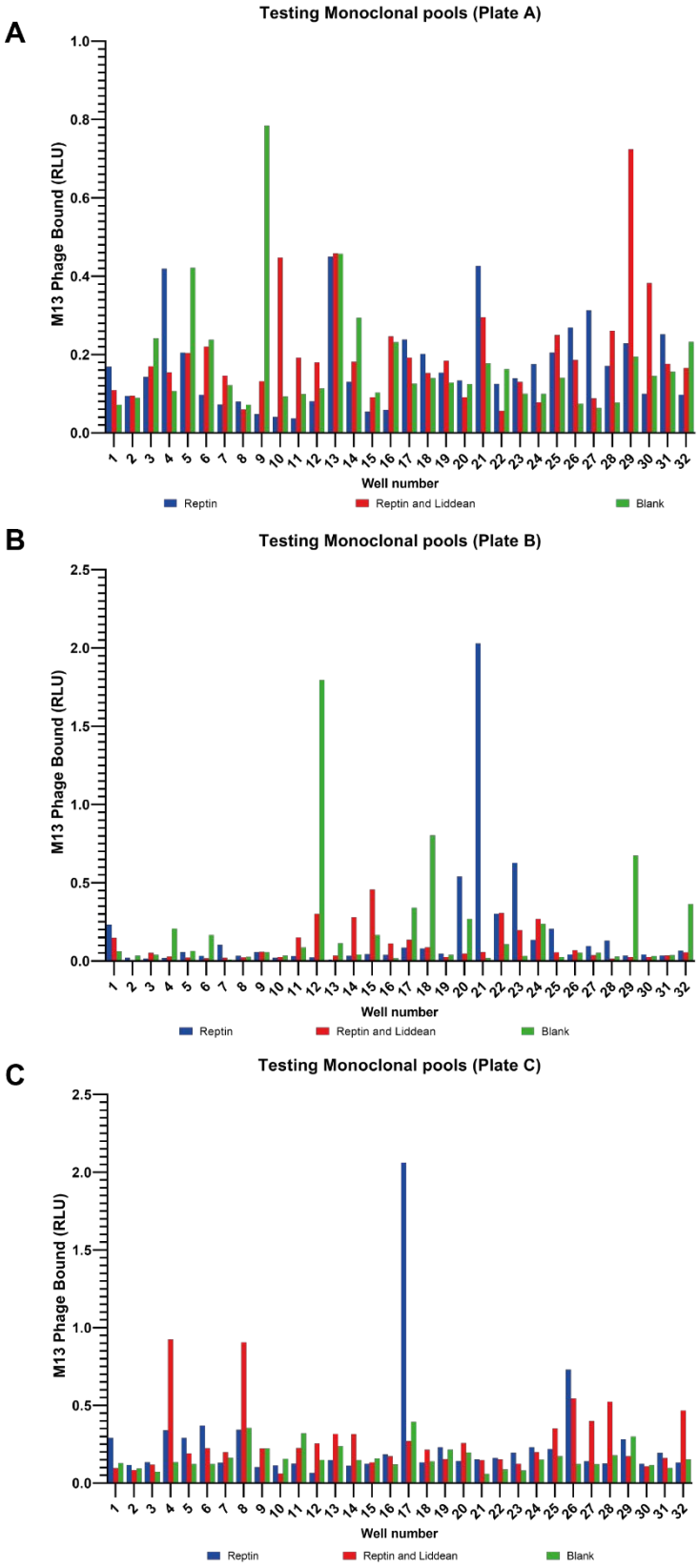


Figure 4.6 Testing Monoclonal Phage Pools for Specific Binding
Monoclonal phage developed from pools binding to Reptin in presence and absence of Liddean

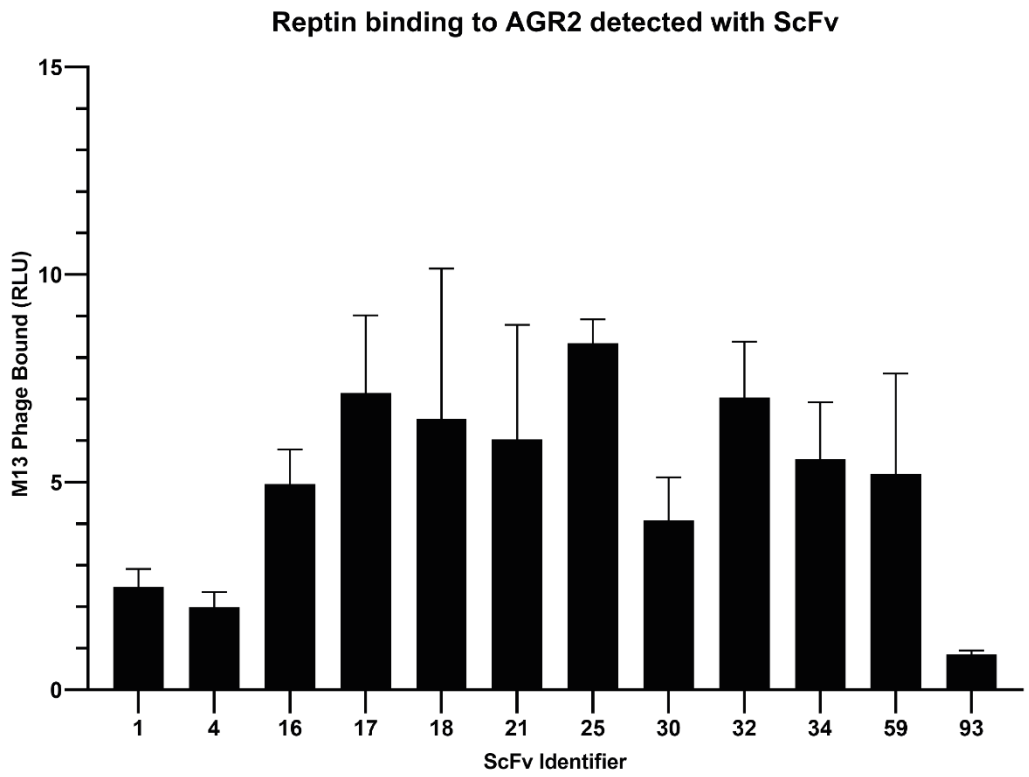


Figure 4.7 Using monoclonal phage against Reptin to identify Reptin bound to AGR2.

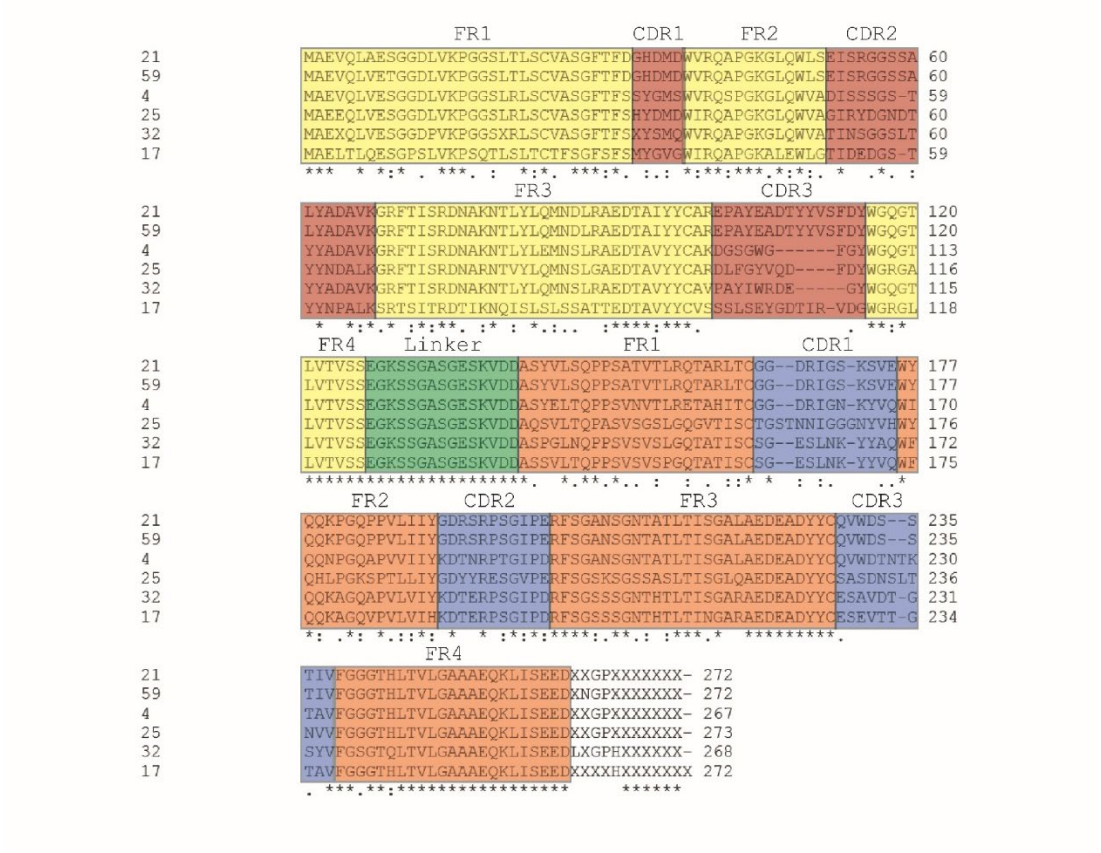


Figure 4.8 Sequences of monoclonal phage aligned showing CDR domains which are responsible for binding affinity of ScFv

*Identical amino acids are identified with *, similar amino acids are highlighted with :, weakly similar amino acids are highlighted with .*

required as they would not be found in the same polypeptide chain in the full-length antibody.

4.4.2 Mapping and testing binding of ScFv to Reptin in different states

These monoclonal ScFv were then tested to see where they were binding to the Reptin sequence. The peptide mapping combined overlapping Reptin self-peptides in groups of 5 to create 'regions' of Reptin. The ScFv was then tested to see which region they bound to (figure 4.9A). Here we can see that some of the ScFv do not appear to bind to any of the peptides, their epitope is likely to not be a linear epitope but a structural one, these ScFv could be useful when investigating conformational change. Their epitopes could be mapped using other techniques such as hydrogen-deuterium exchange mass spectrometry.

Several ScFv can be seen to bind to peptides 1-5, this could be due to the sticky nature of peptide 3, as seen in Chapter 3 figure 3.4C. This experiment should be repeated omitting peptide 3 from the 1-5 well and a peptide 3 only well would confirm this. ScFvs 4 and 25 can be seen to bind to different regions of Reptin, peptides 21-25 and peptides 31-35 respectively, the level of binding here is much higher than the background seen with the other ScFv. As these two were found to have different binding sites with a strong affinity it was decided to map these binding sites onto the crystal structure of Reptin (figure 4.9B).

Following this, these ScFv were selected for mammalian cell expression in addition to ScFv 21. To achieve this, these sequences were ordered for synthesis, with the addition of a CD20 tag to allow easier identification and purification. The sequences were restriction cloned into pSF-CMV vector for mammalian expression (figure 4.10A). The result of this expression can be seen in figure 4.10B, with ScFvs 21 and 25 expressing well. It was not possible to express ScFv 4 under these conditions

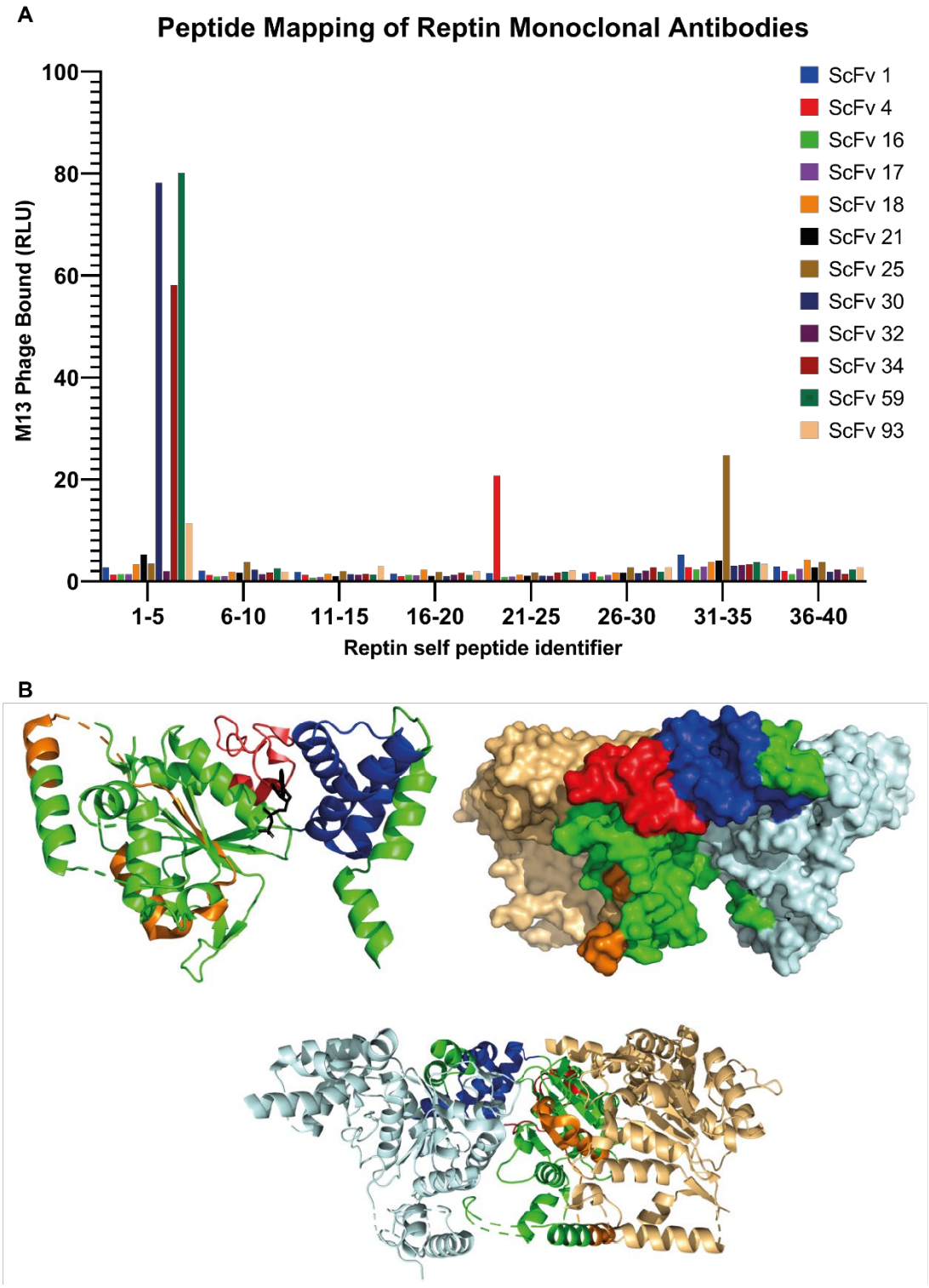


Figure 4.9 Mapping ScFv binding sites on to Reptin.
A) Mapping ELISA to gain information on the epitopes which each ScFv recognize. B) Sites mapped on to 3UK6 structure. The chain in green is a single monomer. ScFv 4 binding site in orange, ScFv 25 binding site in blue and several ScFv binding site in red. Adjacent monomers can be seen in beige and pale blue.

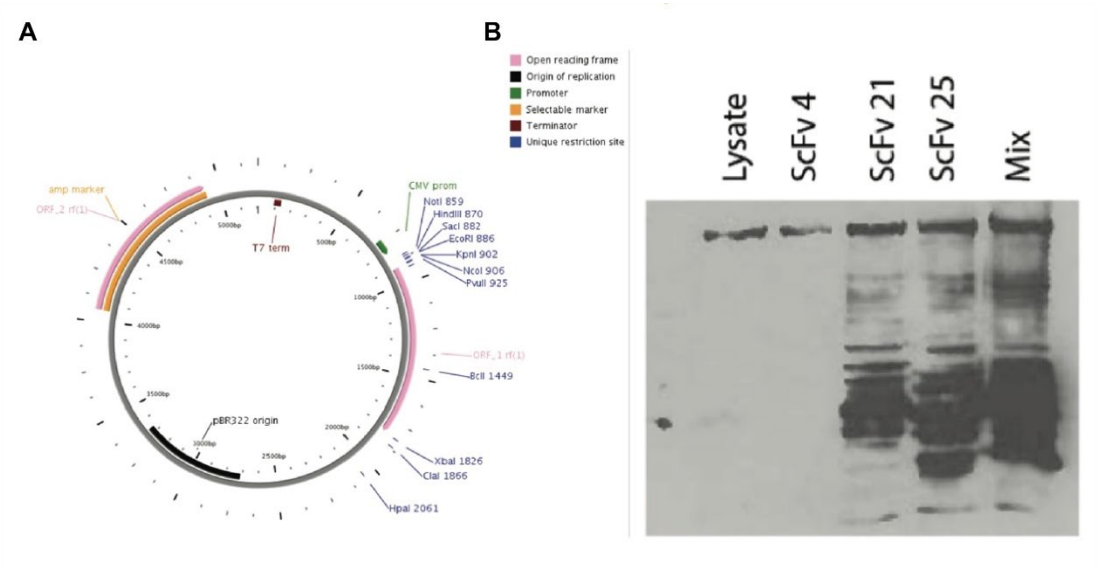


Figure 4.10 Expressing ScFv in Bacteria.
A) Map of plasmid used to express ScFv in bacteria B) Western blot of purified ScFv grown in bacteria.

and more optimisation is required however the sequence of the ScFv 4 vector was seen to have no mutations.

Unfortunately, due to time constraints and techniques still being developed in the Hupp laboratory, it was not possible at the time to carry on with this work, however, the results seen here are very promising that the ScFv developed using this assay would be able to bind different oligomeric forms of Reptin.

4.5 Discussion

Visualisation of the Reptin protein and its binding partners has been achieved in cells using immunofluorescence, from this we can see that both Reptin and its binding partners localise to the same regions of the cell.

To confirm the interaction between Reptin and its binding partners' proximity ligations assays were utilised. In this signal is only observed when the two proteins of interest are in very close proximity with each other either directly interacting or as part of the same complex. The interactions between Reptin and p53 and Reptin and Pontin could both be confirmed in vivo using this technique. Here we can see that the addition of Liddean had an effect of the localisation of the interaction between Reptin and Pontin.

Here, the development of 5 individual single-chain variable-fragments (ScFv) which bind Reptin protein has been achieved. Starting with these 5, it would now be possible to create a new Reptin specific library which can be expressed in bacteria, without the need for the bacteriophage or the use of mammals. This library could be further expanded by using the sequences of the 5 ScFv, the different heavy and light chains could be rearranged and reordered. For example, the heavy chain from ScFv 4 could be partners with the light chain from ScFv 21 creating a new single chain. This could then also be reordered so that when it is expressed instead of the fusion protein being heavy-chain – linker – light-chain, we could have light-chain – linker – heavy-chain. This small difference has been observed to change the binding affinity of an ScFv. This would mean we could have a library of up to 50 single-chain variable fragments able to bind Reptin. This Reptin specific library could then undergo testing to look at changes in binding epitopes and affinities. The ability to have several antibodies which can recognise different conformations of Reptin would be very beneficial.

The epitopes were mapped using a peptide binding assay, this showed that there was variation in binding sites. The ScFv which bound to peptides 1-5 could be further investigated for specificity to Reptin, measuring their binding to peptide 3 and peptides 1, 2, 4, and 5 separately. To investigate the further these epitopes were mapped onto the 3D structure of oligomeric Reptin showing that although the epitopes for several ScFv (peptides 1-5, red) and ScFv 25 (peptide 31-35, blue) are both exposed on the surface of hexameric Reptin, the epitope for ScFv 4 (peptides 21-25, orange) is hidden in the hexamer. This then means that this single chain could be used to identify monomeric forms of Reptin in vivo and in vitro.

In the future, these single-chain variable fragments could be the first specifically designed to look at changes in oligomeric state of a protein. The ability to recognise and localise the monomeric form of Reptin would be very beneficial as it may be this form of Reptin which is oncogenic. Hexameric Reptin is well characterised due to its high stability within the cell, especially its interactions with Pontin, however, as we have previously seen different binding affinities with Reptin for p53 and AGR2 in the presence of ligand, this too could be the state in cells. Combining the Reptin specific ScFv library and proximity ligation assays may give an insight into the amount of the Reptin pool which is involved in each protein-protein interaction. It would also allow an investigation into changes due to drug treatments, stress and stages of the cell cycle.

Due to the size and easy expression of ScFv, they are much more suited for use in a clinical setting that full-length antibodies. The lack of constant regions makes them less immunogenic, as does their expression in bacteria. Nanobodies can also be developed from a single-chain variable fragment by taking either the heavy or light chains. These tend to be much less stable but are more applicable for use in the clinic due to their smaller size and lack of a linker region. It is possible to develop an

immunity against the linker region, targeting the ScFv for destruction by the immune system. Thus, in a clinical setting, ScFv is probably superior to full-length antibodies, however, nanobodies are likely to be the future of this type of therapy due to their smaller size and lack of immunogenic regions.

The use of antibody phage libraries to discover novel biologics is still in development. This technique has revolutionised the field of immunovirology. It is both robust and versatile, as is shown here by being able to develop multiple single chains to a single protein in one experiment. There is an increasing amount of work occurring veterinary field with regards to the use of antibody phage display as, for disease in animals, antibodies made in animals are thought to be considered economically impractical. With the development of phage display techniques, this financial burden is reduced (Bashir and Paeshuyse, 2020). The primary focus has tended to be on use as diagnostics but their potential use in therapeutics exceeds this usefulness. The work carried out here shows the ability of this technique to distinguish between two versions of the same protein. This could be used to distinguish mutations, changes in conformation or oligomerisation or protein interactions which are crucial to disease. Due to being recombinantly expressed they are highly modifiable and could also be used in drug delivery systems with advanced in Click Chemistry (Pisa et al., 2019).

Chapter 5 The impact and significance of ligands on the linear peptide binding activity of Reptin using combinatorial peptide libraries

Some of the work in this chapter was published in 2015 in Chemical Science as a collaboration between the laboratories of Ted Hupp and Nicholas Westwood (Healy et al., 2015).

5.1 Introduction

In Chapters 3 and 4, it has been shown that there are specific changes in protein-protein interactions with the addition of ligands, for example, figure 3.3C and 3.3D where we were examining changes in Reptin binding to p53 and AGR2 respectively when Reptin had previously been incubated with a variety of small molecules.

5.1.1 Combining Peptide Phage Display (PPhD) and Next-Generation Sequencing

Using Peptide Phage Display (PPhD) and Next-Generation Sequencing this change in binding affinity can be investigated further by looking specifically at short peptide motifs (Matochko et al., 2012, Dias-Neto et al., 2009). The PPhD technique uses a commercially available library of peptides which are expressed on the protein coat of bacteriophage, this is available with 7mer and 12mer peptides from NEB. The peptide expressing bacteriophage are exposed to protein bound to a well or tube in the same manner as other interaction assays, this is known as the panning steps. The bound bacteriophage is collected and amplified for subsequent rounds of panning to enrich for strongest binding peptides. In traditional experiments, the specific bacteriophage is selected by plating of eluted phage and sequenced to give a comparatively low number of binding peptides, normally between 100 and 1000 peptides are sequenced from the third round of panning, with less than 100 being

unique (Scott and Smith, 1990). In our protocol we extract the DNA from the whole pool of eluted phage, amplifying the fragments which correspond to the peptide expressed on the protein coat, allowing for millions of sequences to be obtained.

5.1.2 The importance of short linear motifs in protein-protein interactions

Discovery of new protein-protein interactions is a complex process as many of these interactions are highly transient. Many protein-protein interactions appear to use motifs to allow one protein to have many binding partners. These specific, short motifs are linear and do not require a three-dimensional organisation for their recognition. These short motif sequences allow some level of variability between the sequence in the different proteins but the overall pattern to remain the same, these are known as consensus sequences. This means that several proteins can contain the same short, conserved motif, which is used by their binding partner to recognise and target the protein interaction, this is very common for phosphorylation where each kinase is thought to have a specific sequence on its target which pinpoints the location of the addition of a phosphate group (Aitken, 1999).

The importance of these short linear motifs and their flanking amino acids was observed in Chapter 3 with regards to Reptin protein binding to the overlapping peptide library created from the amino acid sequence of Reptin. In figure 3.4B we saw that although peptides 27, 28 and 29 (APVLIMATNRGITRI, GITRIRGTSYQSPHG and QSPHGIPIDLLDRLL) were all overlapping there was a large difference in the binding. There was a large increase in binding between peptide 27 and 28, showing the effect of accessibility of a particular sequence of amino acids. This high binding was completely abolished with peptide 29 showing that a motif of interest must have been contained within the overlap between peptides 27 and 28. During this Chapter, we also saw, in figure 3.5 that even very

strong binding can be attenuated when the protein of interest is incubated with a ligand.

5.1.3 **Peptide Phage Display**

The concept of peptide phage display was first published in 1985, looking at expressing specific amino acid sequences on the protein coat of the phage, P3 (Smith, 1985). From this initial idea, there were several groups which showed that it was possible to express random peptides on the protein coat on P3 creating the first peptide phage display libraries (Scott and Smith, 1990, Devlin et al., 1990, Cwirla et al., 1990). In the following years, many developments were using different phages as the carrier for the library. As mentioned above it is now possible to obtain ready-made libraries commercially.

Using the peptide phage display technology, I performed and experimented comparing apo Reptin, Reptin with ADP and Reptin with Liddean. I then used next-generation sequencing to identify peptides which were bound to the different conditions. The computational analysis allowed in-depth investigation into the large datasets. From this, it was clear the particular motifs were conformationally dependant whereas others are not and show strong binding independent of ligand presence.

5.2 Aims

The experiments in this chapter aim to focus on changes in protein interactions and linear peptide-binding motifs of Reptin when it is exposed to ligands.

We want to investigate:

- The impact of ligand binding on linear peptide-binding motifs
- If peptide phage display can be sued to find or confirm protein-protein interactions
- If linear peptide-binding motifs can be used to identify protein-protein interaction sites

5.3 Carrying out peptide phage display

A schematic of the panning process can be seen in figure 5.1A. In brief, Reptin has bound to the well overnight either alone, or in the presence of ligands. The library of phage was added to the well containing the protein of interest and incubated for 1 hour. Unbound phage was washed away and then the bound phage is eluted. The eluted phage is amplified and then used for a second round of panning in the place of the library. Once completed, the sequenced of the phage are extracted and bar-coded by PCR to each round and condition, a schematic for the location within the phage genome of the expressed peptide in the protein sequence and the primers used can be found in figure 5.1B. The PCR fragments, which can be seen in figure 5.1C, were extracted from the gel and sent for Next-Generation Sequencing. The library was also sequenced to ensure diversity from the offset.

The experiment was carried out twice as the sequencing of the library itself showed a lack of diversity on the first screen. The library in the second screen was seen to have a might higher diversity and the data from this screen was examined in much more detail.

5.3.1 Analysis of the first peptide screen

Even though the library was seen to have a lack of diversity it was decided that there still may be the opportunity to find linear binding motifs from this screen. As next-generation sequencing was used, a large number of peptides were still obtained despite the reduction in the variety of peptides expressed on the coat of the phage. Previously published data used libraries that were not subjected to next-generation sequencing and therefore their diversity was not unknow, due to this it was decided to still probe this data for possible linear peptide motifs.

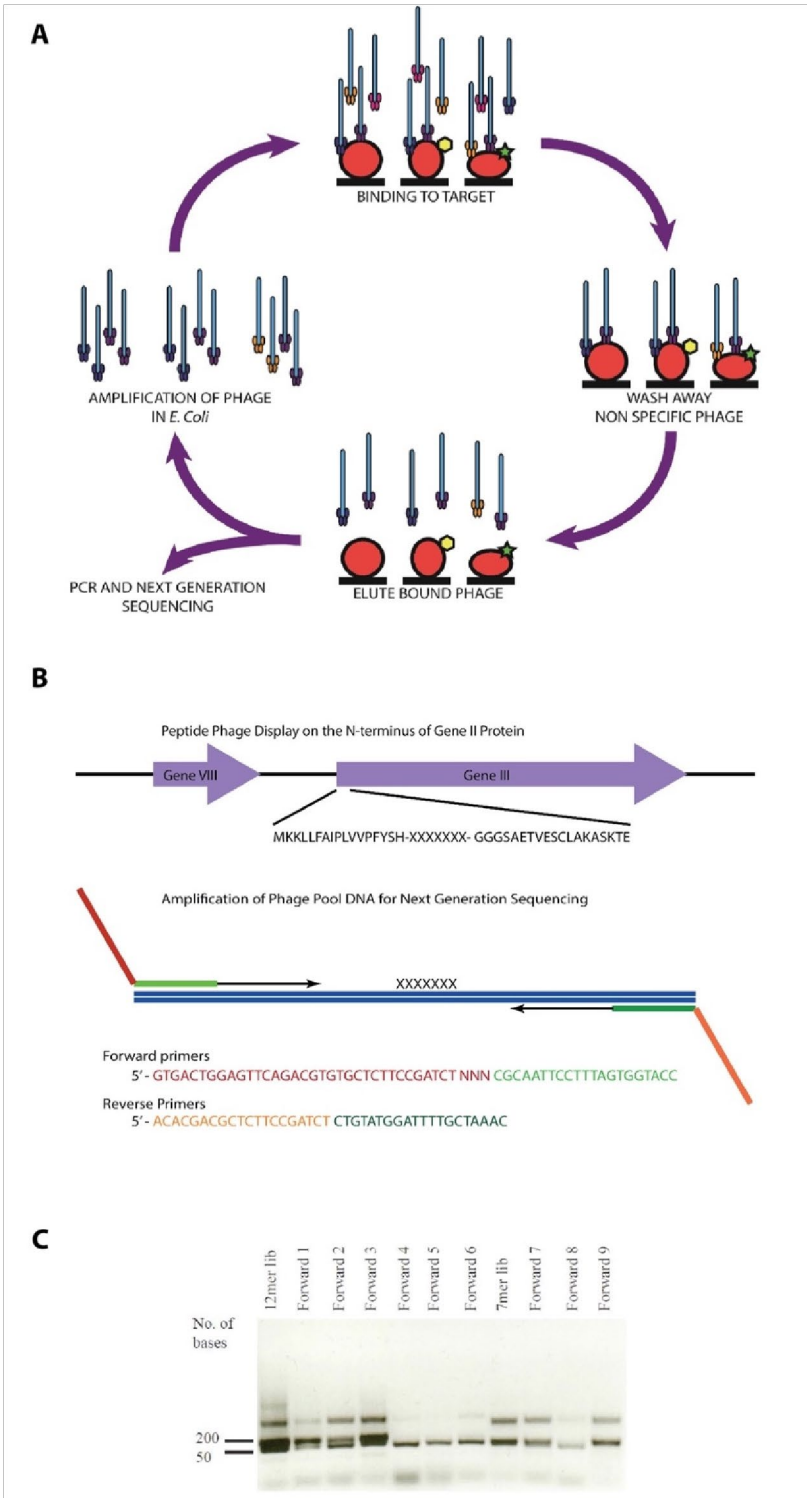


Figure 5.1 Peptide phage display methodology

A) Cycle of panning showing rounds of amplification and elution used to get specific binding peptides. Different expressing peptides on the protein coat is shown as different coloured circles. The protein of interest is seen as a red ball, bound to different ligands (yellow hexagon and green star) resulting in different conformations of the protein of interest. B) Schematic for the location within the phage genome (N-terminal of gene III protein) of the expressed peptide in the protein sequence, XXXXXXX represent the 7-mer peptide library. Amplification of the phage pool DNA ready for next generation sequencing, the primers used are shown here with the barcodes for different samples shown as NNN. C) Representative DNA gel visualising the bands cut for DNA extraction before being sent for next generation sequences. The band of interest can be seen between the 50bp and 200bp markers. With contaminant DNA bands also visible.

analysis and a selection of motifs found can be seen in figure 5.2, all these motifs have a p-value greater than 0.05. We could use this list of motifs to obtain possible binding proteins which could be further investigated. All of the motifs shown here are found to be in peptides found in the Reptin + Liddean samples compared to Reptin only samples.

One motif of particular interest, found from the peptides obtained in the Reptin bound to Ligand samples, is an RER motif (figure 5.3A). This motif is found within the sequence of p53, and more importantly than that, can be found one of the peptides of p53 which had high levels of binding to Reptin (Chapter 3, figure 3.4B). As previously described, the addition of ligand results in an increase in Reptin to binding to p53 protein (Chapter 3, figure 3.3C), so the appearance of this motif is especially encouraging. This p53 peptide of interest is known as peptide 38 (YFTLQIRGRERFEMF) which can be seen in figure 5.3B with the location of RER highlighted. To investigate the importance of this linear motif in the binding of Reptin and p53, a peptide screen was obtained containing variants of peptide 38, firstly an alanine scan of the whole peptide and secondly, truncations and both the N and C termini to observe how binding was affected, this library of peptides can be seen in figure 5.3C.

The protein-peptide assay, seen in figure 5.3D shows the importance of the presence of the RER motif and how important the accessibility of the motif is. The changes in the length of the peptide showed how its location is so important. The peptide with the highest binding, peptide 24 (YFTLQIRGRER), is the one which contains the motif at the C-terminus making it the most freely accessible. We can also see how disruption of motif or its adjacent amino acids can affect levels of binding. Peptides 6 (YFTLQ**A**RGRERFEMF), 7 (YFTLQI**A**GRERFEMF), 9

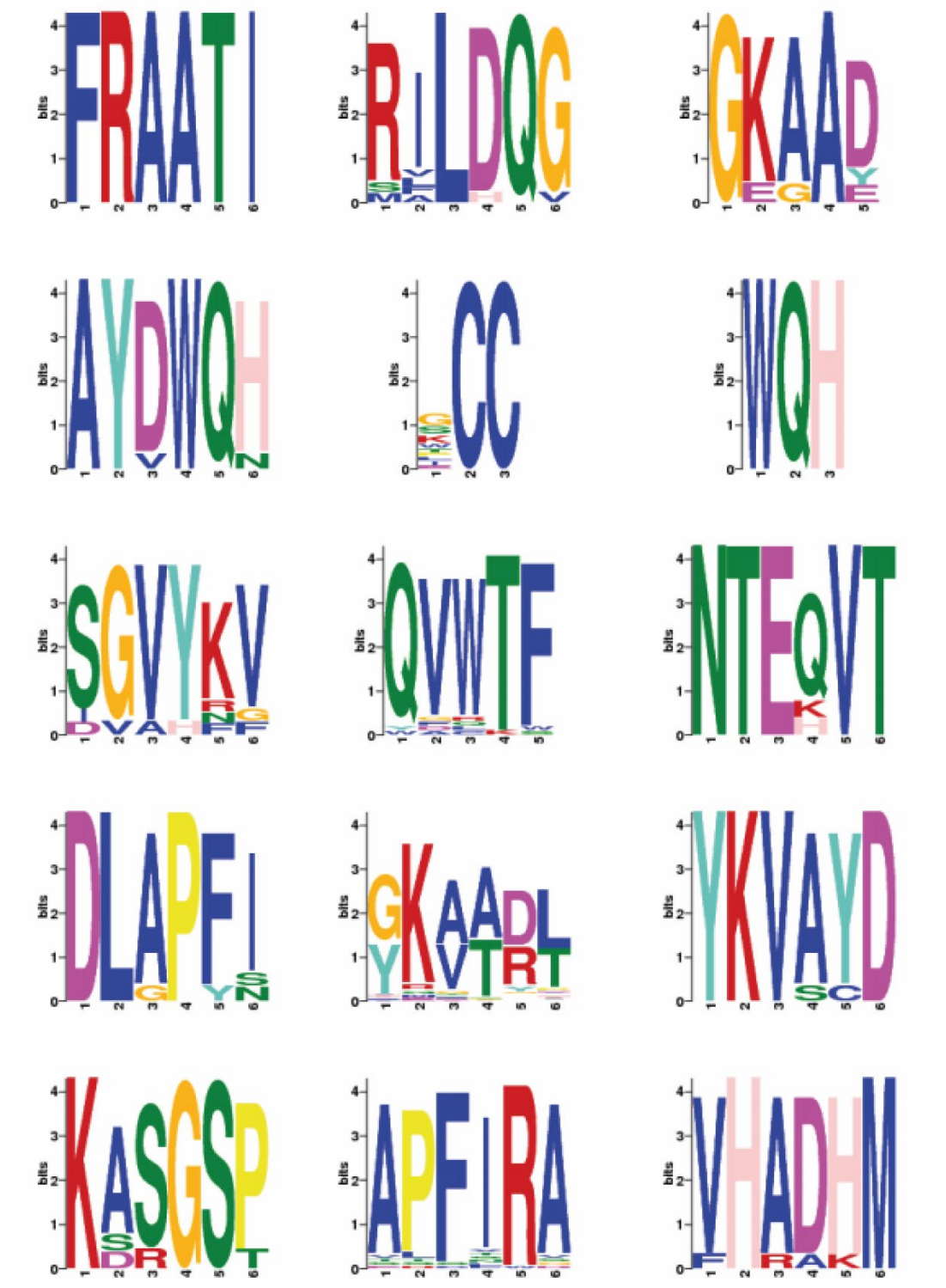


Figure 5.2 Top 15 motifs discovered using first Peptide Ph.D screen
Motifs discovered by subjecting the peptides found in the Reptin + Liddean samples compared to Reptin only samples. All these motifs have a p value greater than 0.05. The software used is MEME SUITE (Bailey et al., 2009).

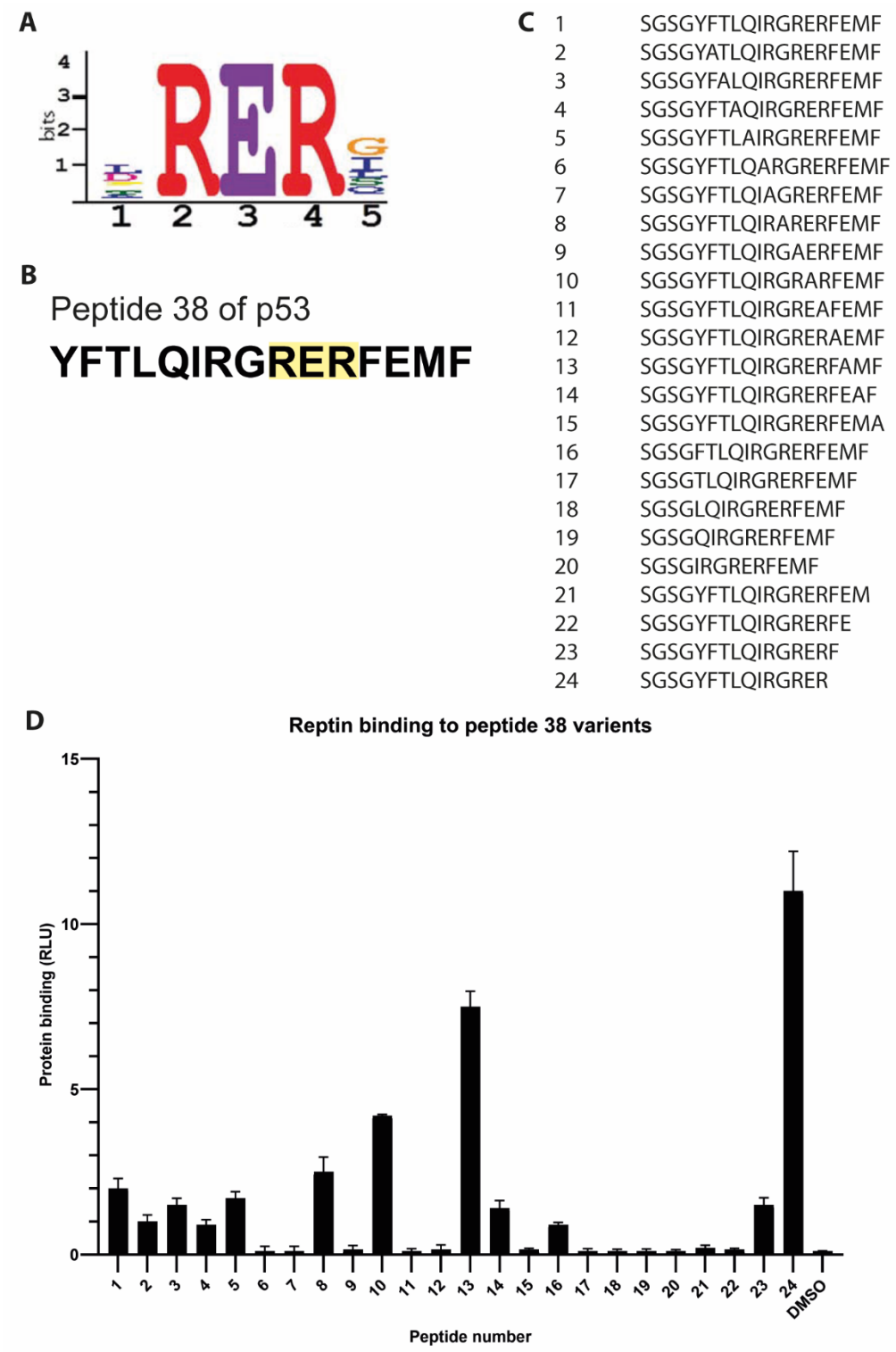


Figure 5.3 RER motif discovered in p53
A) RER motif as given by MEME SUITE (Bailey et al., 2009). B) Amino acid sequence of Peptide 38 of p53 with RER motif highlighted. C) Library of peptides developed from peptides 38 including alanine scan and truncations. D) Peptide-protein assay showing Reptin binding (RLU) to peptides in C. (Healy et al., 2015)

see the disruption of either arginine residues (peptides 9 and 11), as well as changes in the amino acids flanking the motif, tend to result in loss of binding. Peptide 8 (YFTLQIR**A**RERFEMF), whose mutation is directly before the RER motif, appears to be unaffected in its binding level. Conversely, peptides 10 (YFTLQIRGR**A**RFEMF) and 13 (YFTLQIRGRERF**A**MF) both appear to have a large increase in binding affinity, this is noteworthy because this shows that this consensus sequence is more likely to be an R-X-R motif as the mutation of the central amino acid resulted in an increased in binding.

The screen shows how changes in a single peptide, either with an alanine scan or shortening of the peptide, can have a drastic effect on the strength of an interaction. These short motifs are often highly specific and are not very flexible with regards to mutation. There are several diseases which are the result of point mutations, for example, sickle cell anaemia is caused by a single nucleotide change resulting in the peptide chain containing valine rather than a glutamic acid, this changes the charge of the amino acid causing a conformational change in the result of the peptide chain (Elion et al., 1992).

5.3.2 **Analysis of second peptide screen**

The sequencing for the library used in this screen was found to be much more diverse than that of the first screen. This allowed for bioinformatic analysis of the next-generation sequencing data from the second screen was carried by Stuart Aitken (IGMM Bioinformatics Core Facility).

Figures 5.4 and 5.5 show the results of this analysis and were created by Stuart Aitken. Figure 5.4A shows the percentage of sequences which were obtained for each of the conditions relative to one another, AAA is the library itself. From this data, it is possible to see that the number of peptides sequences in each condition is

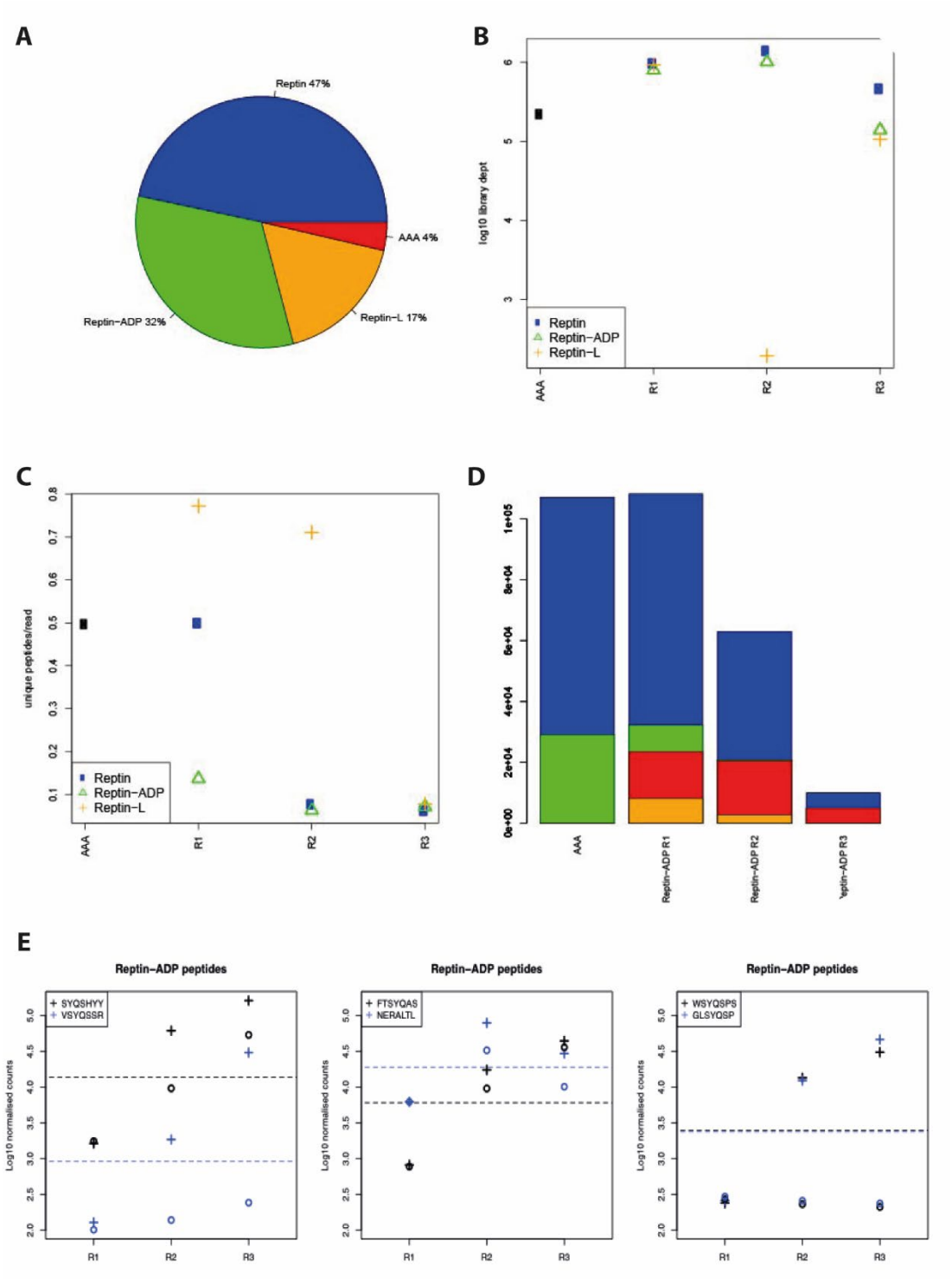


Figure 5.4 Bioinformatic analysis of second peptide phage display screen. *A) Pie chart showing the percentage of sequences which were obtained for each of the conditions relative to one another, AAA is the library, Reptin, Reptin-ADP and Reptin-L(iddean). B) Comparison of the depth of sequence between samples. C) The Number of unique peptides in each of the samples across the rounds, the average of the library is also shown. D) Total number of unique peptides across the different rounds for Reptin-ADP. Peptides which only appear in one round or just the library (blue), peptides found in the subsequent round (green) peptides seen in the previous round (red), peptides found in both earlier and later rounds (orange). E) Normalised read counts for the six most highly expressed peptides in Reptin-ADP. SYQSHYY, VSYQSSR, FTSYQAS, NERALTL, WSYQSPS, GLSYQSP. (Created by Stuart Aitken, University of Edinburgh)*

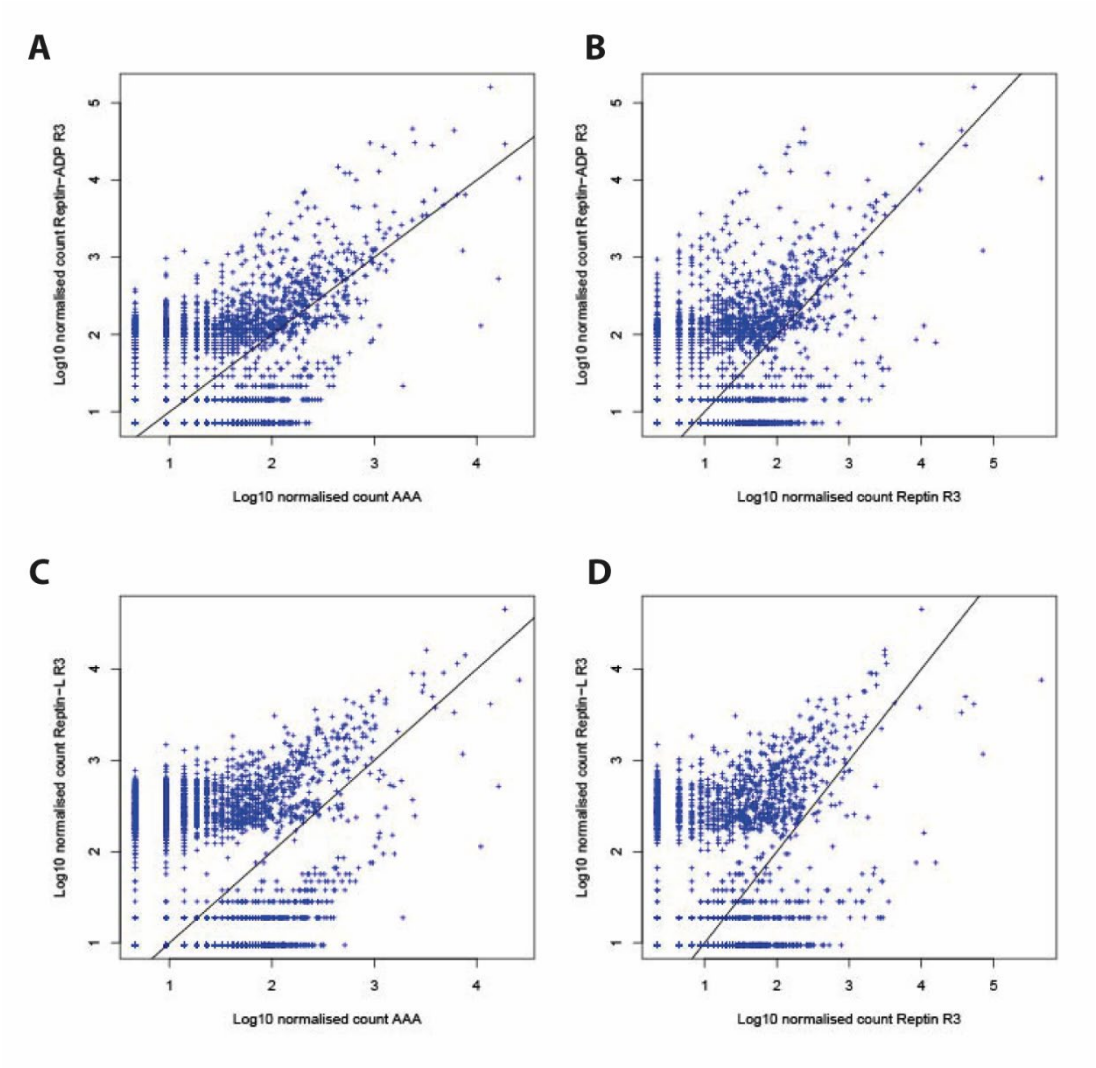


Figure 5.5 Volcano plots comparing peptides seen under different conditions.

A) Library with Reptin-ADP round 3. B) Reptin round 3 with Reptin-ADP round 3. C) Library with Reptin-Liddean round 3. C) Reptin round 3 with Reptin-Liddean round 3. Diagonal lines represent the average correlation between the two samples. (Created by Stuart Aitken, University of Edinburgh)

round of panning, except the Reptin-Liddean sample from round 2 (Figure 5.4B), the reduction in the size of this sample explains the reduction in the number of sequences overall for all rounds, seen in the pie chart in figure 5.4A, in comparison to the other experimental conditions.

The number of unique peptides in each of the samples is probed in figures 5.4C and 5.4D. In figure 5.4C we see the Reptin-Liddean samples (yellow cross) appear to have many more unique peptides than the others, this is opposing with the other ligand-bound form of Reptin, Reptin-ADP, which appears to have the least number of unique hits. The large number of unique peptides seen in all conditions means that there are specific binding profiles for each of the forms of Reptin. The bar chart in figure 5.4D shows the total number of unique peptides across the different rounds, this time looking specifically at Reptin-ADP. In blue we have highlighted those which only appear in one round (or just the library), those that are found in the subsequent round in green, those that have been seen in the previous round in red and those that are found in both earlier and later rounds in orange. Here we can observe that many of the peptides are only sequenced in a single round, however, the number of unique peptides decreases over the rounds, showing an increase in specificity as we perform the rounds of panning. This analysis was not possible for the Reptin-Liddean samples due to the lack of data from the second round of panning.

Normalised read counts for the six most highly expressed peptides in Reptin-ADP are found in figure 5.4E and are shown by crosses. The circles show the number of reading counts from the Reptin only samples for the same peptides and the parallel lines represent the levels observed for these particular peptides in the library. Their sequences are as follows - SYQSHYY, VSYQSSR, FTSYQAS, NERALTL, WSYQSPS, GLSYQSP. We can see follows how the read counts are amplified for

each peptide throughout the rounds showing the increase in specificity if these are true binding peptides of Reptin.

The volcano plots in figure 5.5 compare the results from the different conditions to the library itself and the Reptin conditions. The conditions compared were the library and the Reptin-ADP round 3 sample (figure 5.5A), the Reptin sample with the Reptin-ADP sample, both from round 3 (figure 5.5B), the library and the Reptin-Liddean round 3 sample (figure 5.5C), the Reptin sample with the Reptin-Liddean sample, both from round 3 (figure 5.5D). Due to the shape of these plots, it is possible to suggest that we do see a combination of peptides which are binding under all conditions in addition to those which are specific to each condition (outlying points). The diagonal lines represent the average correlation between the two samples.

Peptides were ordered for commercial synthesis to see if the amino acid sequences observed to have variation in their binding to the different forms of Reptin in the screen could also do so in peptide-protein binding assays, in a similar manner to those performed in Chapter 3, figure 3.4. Figure 5.6 shows many different peptides binding to Reptin in the absence and presence of Liddean. H6, a peptide from AGR2 which shows the reduction in binding on the addition of Liddean, is used as a control on both sets of samples (peptide 1). Figure 5.6A shows peptides selected as binding to Reptin in the absence of ligand, these do a variation, some are seen to have a reduction in binding in the presence of ligand whereas others can be seen to have increased binding. In contrast, those peptides selected as Reptin-Liddean binders, in figure 5.6B, do consistently show an increased affinity for Reptin in the presence of ligand.

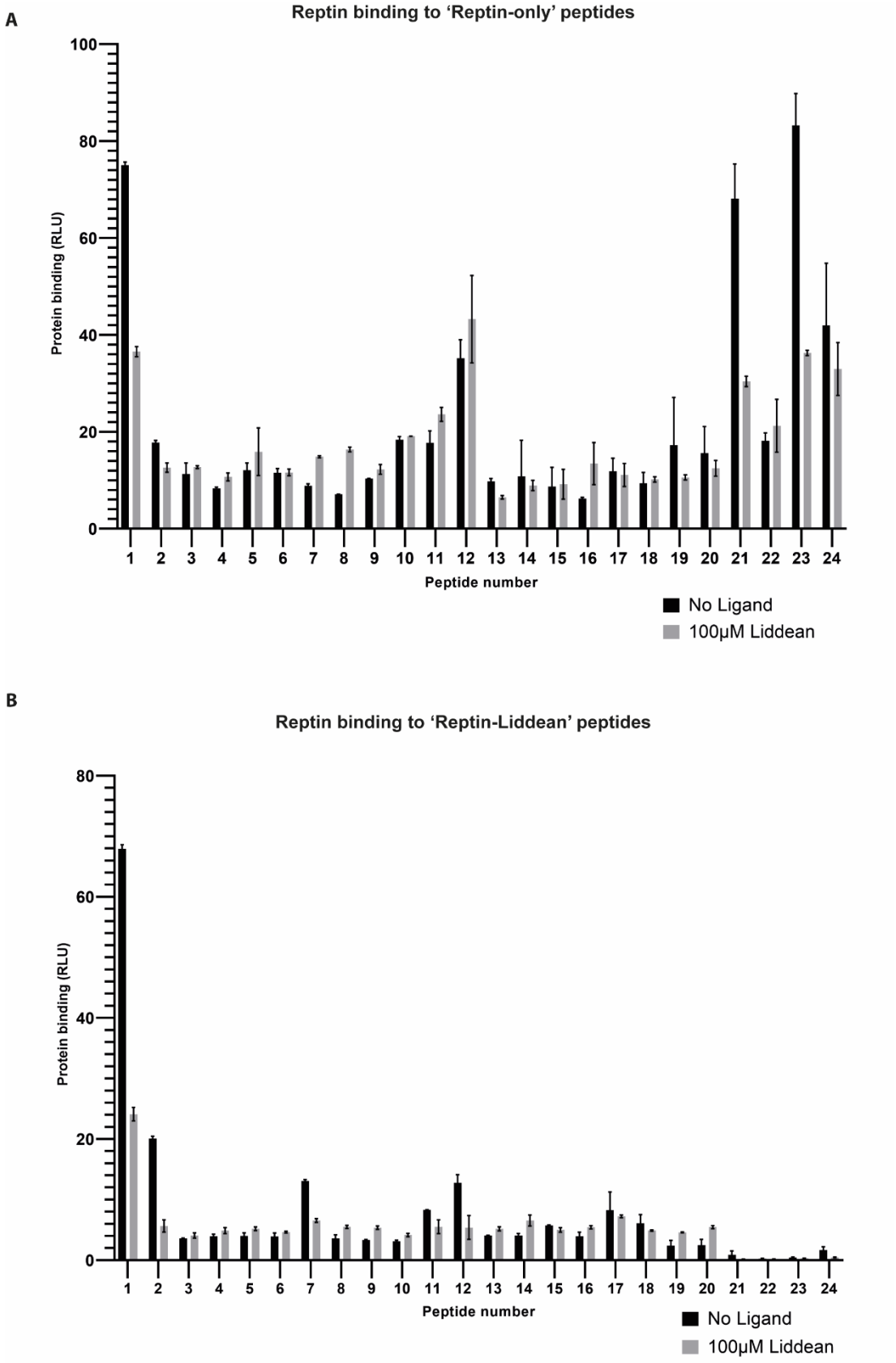


Figure 5.6 Reptin binding to synthesized peptide phage display peptides in the absence and presence of Reptin.
A) Peptides designated Reptin only binders. B) Peptides designated as Reptin-Liddean binders.

5.4 Discussion

Combining peptide phage display and next-generation sequencing allowed for a much deeper screen of peptides which can bind to Reptin than would have been possible with the traditional method of sequencing individual bacterial colonies infected with phage.

In the first screen, we were able to identify several binding motifs, one of which, RER (figure 5.3A), was then used to identify a possible binding interface between Reptin and p53. Following on from the studies in Chapter 3, I already knew that these two proteins were able to bind one another in both the protein-protein and peptide-protein interaction assays. Using an overlapping peptide library from the p53 sequence we identified two peptides from p53, peptide 31 (RNSFEVRVCACPGRD) and peptide 38 (YFTLQIRGRERFEMF) which had a high binding affinity for Reptin. When comparing these two peptides to the results of this screen we identified the presence of the RER motif in peptide 38 (YFTLQIRG**RER**FEMF).

To enable confirmation of the importance of this peptide and motif a peptide library was created for peptide 38 containing an alanine scan and truncated peptides (figure 5.3C). This assay showed the importance of this site in the binding interface between Reptin and peptide 38. Also, it revealed the possibility of a more complex binding motif at this site.

Due to the greater diversity in the library used in the second screen, it was possible to perform a more complex bioinformatic analysis of the data. This analysis enabled confirmation of specific pools of peptides which can bind to Reptin under different conditions (figure 5.4). We can also observe how specific peptides are enriched for binding throughout the experiment.

From the bioinformatic analysis, several peptides were selected as binding to Reptin in a single condition. These peptides were then synthesized for testing in the same manner as is chapter 3. From this assay, there is a diversity of results from the different peptides showing a landscape of Reptin binding affinities.

In this chapter novel binding peptides have been identified for their ability to bind Reptin. It has also been possible to map a possible binding interface between Reptin and p53. The next step for this investigation would be to the first model and then mutate the residues IRGRERFE on the binding of Reptin protein to peptide 38 in the full length p53 protein and observe the changes in protein-protein interaction. If the results seen in the protein-protein assays mirrored the results from the peptide-protein assays, then it is likely that these residues on p53 are the binding site for Reptin.

Reptin is known for interacting with many chromatin and chaperonin signalling proteins. The techniques used here provided novel insights into the protein-protein interactions of Reptin. With further, more detailed analysis, it may be possible to create an entire Reptin protein-protein interaction network, using the motifs discovered to identify novel binding partners, this has previously been done using phosphorylation dependant signalling (Miller et al., 2008). The small linear peptide motifs can form dynamic and highly specific docking sites on one protein for another protein (Van Roey et al., 2014, Tompa et al., 2014). Small molecule stabilisation and destabilisation, as we see here, provides a promising approach towards the modulation of protein functions on a much larger scale (Jubb et al., 2012). An example of this is a linear peptide motif which binds into a groove on the surface of MDM2 has been developed for use in the clinic as an MDM2 inhibitor, allowing activation of the p53 pathway to decrease proliferation in tumours (Ramasamy et al., 2014).

Small linear peptide motifs form a vast and largely untapped landscape in both scientific research and the clinic (Basse et al., 2013). Approaches focusing of the importance of these small linear peptide motifs provide new possibilities for drug discovery (Ramasamy et al., 2014, Kumar et al., 2020) and there is emerging information that proteins, such as Reptin, which have a large number of binding partners, may exhibit features making them linear motif binding hub proteins able to help regulate the large protein complexes which they form part of (Jespersen and Barbar, 2020).

Chapter 6 Examining changes in protein oligomerisation and conformation with the addition of ligands

Much of the work in this chapter was published in 2019 in Journal of Proteomics as a collaboration between the laboratories of Ted Hupp, Borek Vojtesek and Douglas Houston (Remnant et al., 2019). Some of the work in this chapter was published in 2015 in Chemical Science as a collaboration between the laboratories of Ted Hupp and Nicholas Westwood (Healy et al., 2015).

6.1 Introduction

The super-family of ATPases associated with various cellular activities (AAA+) are proteins which all contain a nucleotide-binding domain, the AAA domain, which facilitates ligand binding (Snider et al., 2008). It is well documented that members of the AAA+ protein family can change conformation when bound to their ligand. This change in conformation could mean a change in the protein-protein interactions a particular protein takes part in. This could result in changes in the role a particular protein plays within the cell and the pathways it is involved in (Weikl and Paul, 2014, Keskin, 2007).

6.1.1 3D structure of Reptin and Pontin

Reptin and its homolog Pontin form a heterododecamer. This complex was examined using cryo-EM (Lopez-Perrote et al., 2012), in the presence and absence of ligand, in this case, ADP and two derivatives. Two conformations have been observed in both the apo and ligand-bound complexes. These two conformations of the human proteins, a compact and a stretched conformation, are thought to coexist to allow regulation of the accessibility of the DNA binding domains. The appearance of compact and stretched conformations has previously been observed in yeast proteins, and therefore, as Reptin and Pontin are derivatives of yeast RuvB, these

conformations can likely be found in vivo as they show so much similarity with RuvB itself. These two conformations are thought to be caused by a change in domain II of the proteins, the domain which has previously been removed to allow the formation of crystals for X-ray crystallography experiments, therefore, removing the flexibility of the proteins. If it is this domain is causing two conformations of the protein then removal of this domain can allow for a single conformation and therefore the creation of crystals with proteins of identical conformation as opposed to a mix of two conformations which would not be compatible with the X-ray crystallography technique. The conformational change observed in this experiment is thought to be due to the need for regulation of DNA or RNA binding. Domain II which is thought to contain the nucleotide-binding domain, however, this has not been experimentally proven and it only assumed due to the presence of the binding motif (Lopez-Perrote et al., 2012).

6.1.2 **Techniques for measuring conformational change**

It is possible to observe of changes in conformation through many different biophysical techniques such as hydrogen-deuterium exchange mass spectrometry (Katta and Chait, 1993), ion mobility mass spectrometry (McCullough et al., 2008), partial digestion with proteases and native gels.

6.1.2.1 Hydrogen-deuterium exchange mass spectrometry

Hydrogen-deuterium (HDX) mass spectrometry (MS) (HDX-MS) (Katta and Chait, 1993) is a biophysical technique which allows observation of conformational changes of a protein or proteins by comparing the amount of deuterium present in a peptide before and after specific treatment or interaction. The way this works is that the hydrogens exposed to the outside of the proteins exchange rapidly for deuterium (“heavy hydrogen”) when they are. Immersed in deuterated/heavy water. The longer the protein or proteins of interest are immersed in deuterated water the more

deuterated the protein becomes with hydrogen atoms found to be less exposed slowly changing over time. This means that use of a time course of exposure can give some idea on the overall globular structure of a protein and if this is changing with the addition of a binding partner or ligand. A schematic for this process can be found in figure 6.1B, here the hydrogen is shown in blue with deuterium marked in blue. To stop the exchange both the pH of the solution and the temperature must be dropped, this locks the deuterium into the peptide backbone of the protein. For mass spectrometry analysis of these proteins of interest, the samples are then digested with trypsin (or another well-characterised protease) before being loaded on to the mass spectrometer. With peptide sequence in the mass spectrometer, it is possible to follow the amount of deuteration on a specific peptide throughout the experiment without the need for further labelling with SILAC (Stable Isotope Labelling by Amino Acids in Cell Culture (Ong et al., 2002)) or TMT (Tandem Mass Tag (Thompson et al., 2003)) however for complex mixtures (such as multiple proteins) the addition of these steps can be beneficial.

6.1.3 **Methods for observing changes in oligomeric state**

There are many methods for observation of the oligomeric state of a protein. As spoken about above, hydrogen-deuterium exchange mass spectrometry (Katta and Chait, 1993) can give information on transitions between monomer and multimers as well as an idea of the binding site, (McCullough et al., 2008). Size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALs) (Tarazona and Saiz, 2003) gives information about the exact molecular weight of a complex allowing for interpretation of its oligomeric state.

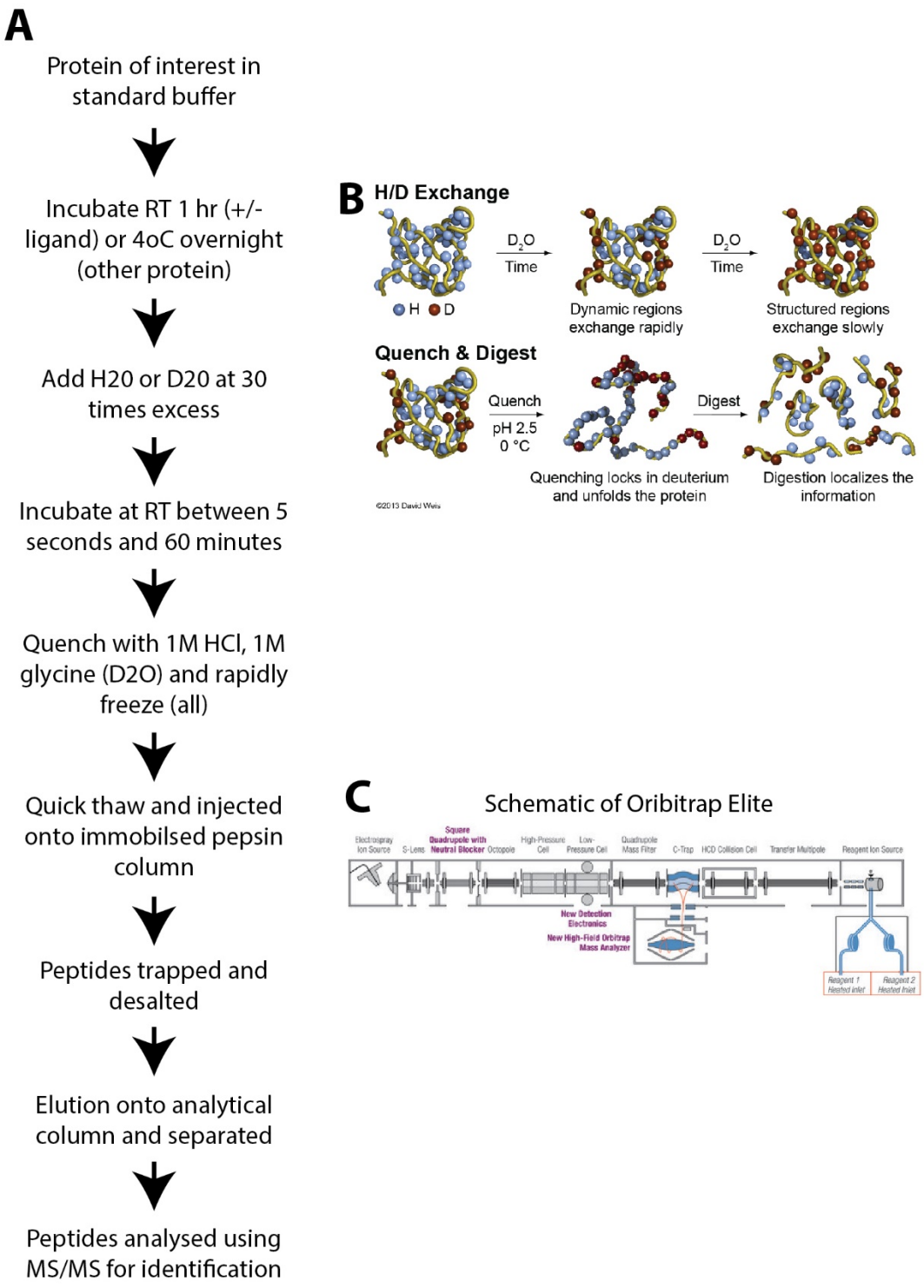


Figure 6.1 Hydrogen-deuterium exchange mass spectrometry (HDX-MS). *A) Flowchart showing protocol for HDX-MS performed here. B) Schematic for how the exchanges of hydrogen atoms for deuterium atoms along the peptide backbone occur over time followed by how this is quenched to lock the deuterium into the peptide backbone before digestion and then analysis by mass spectrometry. C) Schematic of Orbitrap Elite the mass spectrometer used in this study.*

Gel filtration chromatography is used commonly to separate proteins and protein complexes of different sizes. Larger complexes are eluted from the column first while smaller complexes take longer, this is due to the matrix which is contained within the column. This matrix contains pores of varying size meaning that the largest proteins/complexes travel the quickest as they have a much shorter distance to travel compared to smaller proteins/complexes which can enter many channels within the matrix taking longer for them to elute from the column (Gai et al., 2004). This technique is commonly used to separate complex mixtures of proteins and is the first step in many mass spectrometry techniques and is the 'SEC' in SEC-MALS.

Analytical ultracentrifugation functions in a similar manner to gel filtration, separating protein/complexes based on their physical attributes (Cole et al., 2008). Larger proteins/complexes will be found lower in the column whilst smaller proteins/complexes will be found more towards the top. The column is made by preparing a gradient, usually sucrose, and then spinning the sample at high speed in a vacuum for serval hours. This results in a distribution of the proteins/complexes after equilibrium is reached, that is directly related to the size of the protein/complex. This is an equilibrium sedimentation experiment.

6.1.4 **Methods for examining the ATP potential of a protein**

As Reptin is a member of the AAA+ protein family with a known ATP binding site it seems logical that Reptin would be able to bind to ATP and hydrolyse it to ADP. There are several techniques which are commonly used to look at ATP binding and ATPase activity including measuring the amount of radioactively labelled ATP bound to a protein of interest and measuring levels of hydrolysis of ATP.

6.2 Aims

The experiments in this chapter aim to observe small ligand binding using hydrogen-deuterium exchange and the impact of this binding on the conformation and oligomerisation of the Reptin protein.

We want to investigate:

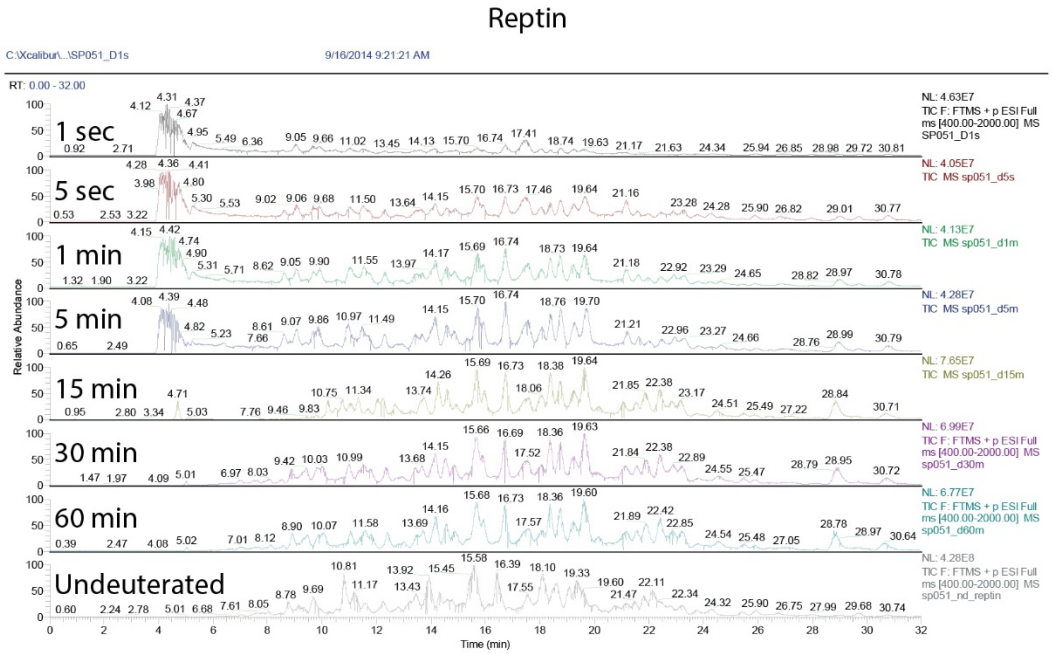
- The ability of Liddean to bind into the ATP binding pocket of Reptin
- Sites of importance in changes in oligomerisation of Reptin
- The effect of point mutations on the conformation and oligomerisation of Reptin
- The role of oligomerisation in ATP binding and ATPase activity of Reptin

6.3 Hydrogen-deuterium mass spectrometry of Reptin in the absence and presence of Liddean

Initially, Reptin was taken in the absence and presence of Liddean, a small molecule ATP mimetic, and exposed to deuterium for various time points between 1 second and 60 minutes to observe the exposed peptides in Reptin and if these changed after addition of a ligand. After the reactions were quenched using 1M HCl with 1M glycine in the case of the deuterated samples, all samples were quickly frozen. These samples were loaded onto an immobilized pepsin column. Peptides were then trapped and desalted on-line on a peptide microtrap and finally eluted onto an analytical column and separated by a linear gradient. These peptides entered the Orbitrap Elite (schematic available in figure 6.1C) using electrospray ionisation. Tandem mass spectrometry was used to obtain identification of specific peptides and then the heights of the peaks lacking deuteration and corresponding deuterated peaks for each peptide were compared to obtain information of the fold change in deuteration in the two peptides. A workflow can be found in figure 6.1A.

The results from the injections of the different samples into the analytical column can be seen in figure 6.2. Here we can see the changes in retention times of the peptides of Reptin, relative to the time they are incubated in deuterated water. In panel A, the Reptin only samples, we see the introduction of several peaks in the 1 second to 5-minute incubated samples compared to the protein lacking deuteration. These peaks are lost in the longer incubations with an introduction in peaks with a longer retention time at around 29 to 31 minutes which are not seen in the samples which have not been exposed to deuterated water. However, panel B, Reptin with Liddean, lacks the large number of peaks found to elute from the column between 4 and 5 minutes but the peaks seen around 29 to 31 minutes are present in all samples except the sample which were not exposed to deuterated water.

A



B

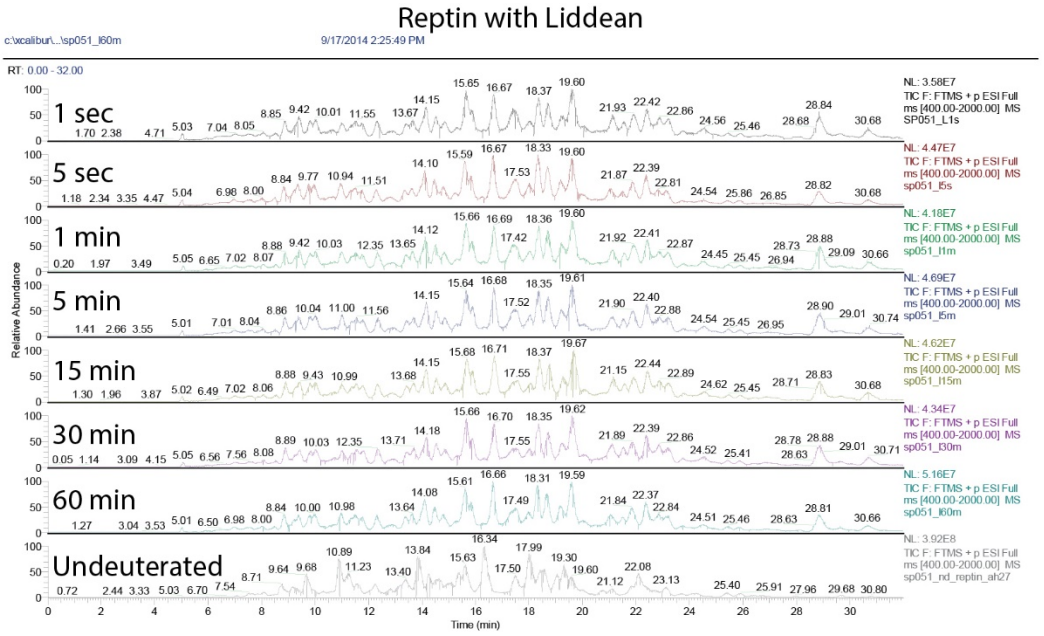


Figure 6.2 Elution profiles of Reptin in the absence and presence of Liddean showing changes in profiles over the deuterium time course, 1s.
A) Profiles from Reptin alone incubated for 1 hour at room temperature before expose to deuterium. B) Profiles from Reptin after incubation with Liddean for 1 hour at room temperature before expose to deuterium

Due to the techniques used in mass spectrometry of ionising the peptides to allow them to be visualised by the mass spectrometer, it is not possible to 'see' all the peptides of a protein, in most cases around 60 % of the protein is seen, with regards to Reptin we can see around 90% of the protein, this is much higher than with the majority of proteins making it perfect for this type of analysis. A visualisation of the coverage can be seen in figure 6.3A, showing that only the N-terminus and a section of the C-terminus is lacking in coverage. The changes in deuteration of the peptides of WT Reptin after 30 minutes incubation can be found in figure 6.3B showing that, at this time point, a large amount of the protein has undergone hydrogen-deuterium exchange compared to the undeuterated sample, the higher the deuteration (in red) shows peptides which are exposed on the surface of Reptin whereas the blue/green peptides are those which tend to be more internalised, those peptides which show crossover with colours can show more specificity for the sites of deuteration as the values are averaged across the whole of the peptide.

We can look at changes in specific peptides by observing the spectra from the mass spectrometer which has been given by a single peptide. The mass shifts to the right of the graph show a heavier peptide thus a deuterated peptide. As an example, the spectra for the peptide AQALGPDTPFTAIAGSE can be seen in figure 6.4A with the undeuterated peptide in red with the deuterated peptide (30-minute incubation) in blue. You can see the deuterated peak is shallower and broader and this is due to variation in the amount of deuterium in the peptide. The specific changes observed in the levels of deuteration at 30 minutes comparing apo Reptin to Reptin with Liddean can be seen in figure 6.4B, here a reduction in deuteration is seen in blue/green, this corresponds to the peptide having less exposure to the outside aqueous environment, an increase in deuteration can be seen in yellow, this is

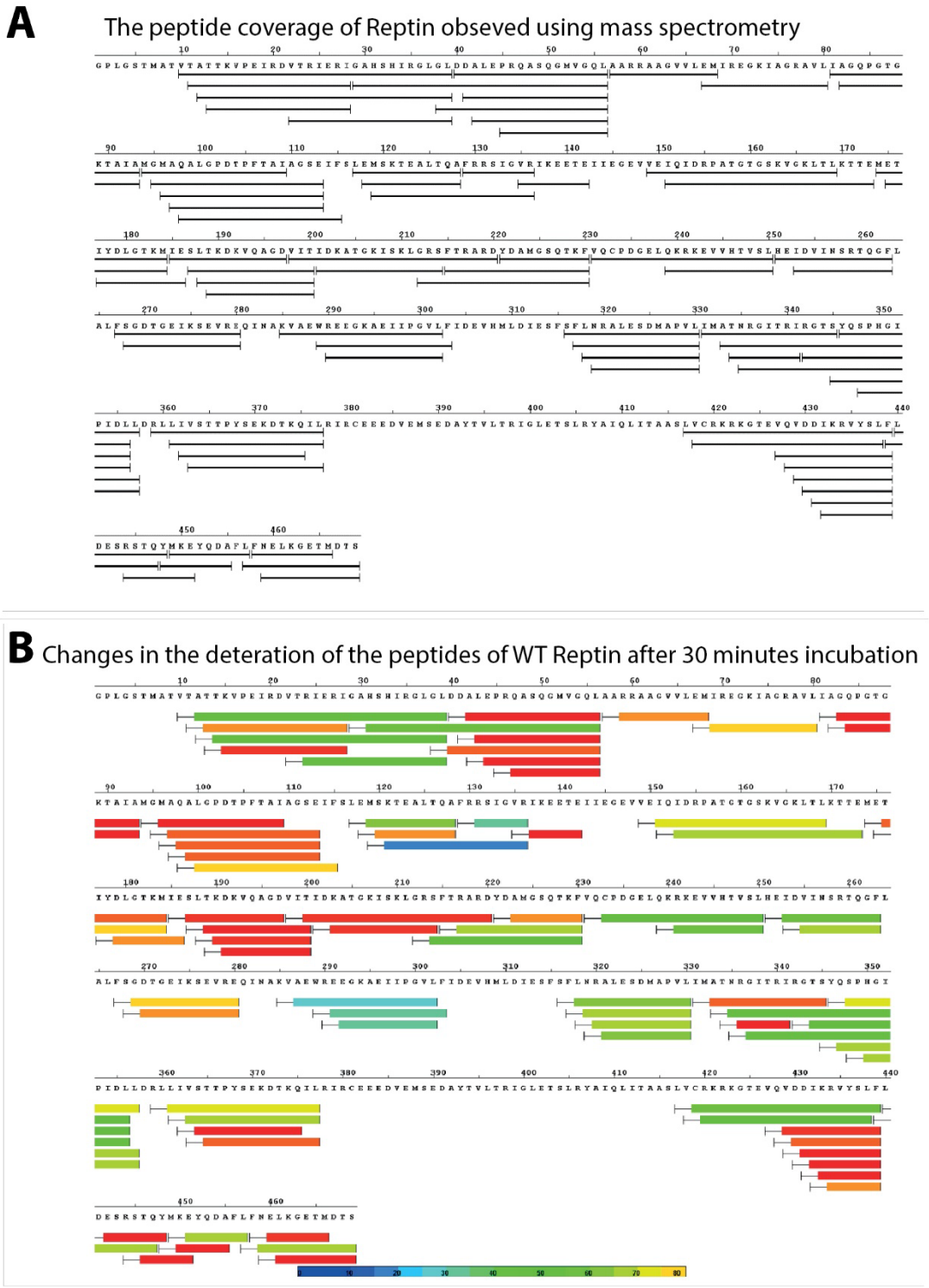


Figure 6.3 Peptide coverage of Reptin.
A) The peptide coverage of Reptin observed by the mass spectrometer, around 90% of the protein is able to be visualised. B) Changes in the deuteriation of WT Reptin peptides observed after 30 minutes incubation with deuterated water, colours range from blue to red showing low to high levels of deuterium.

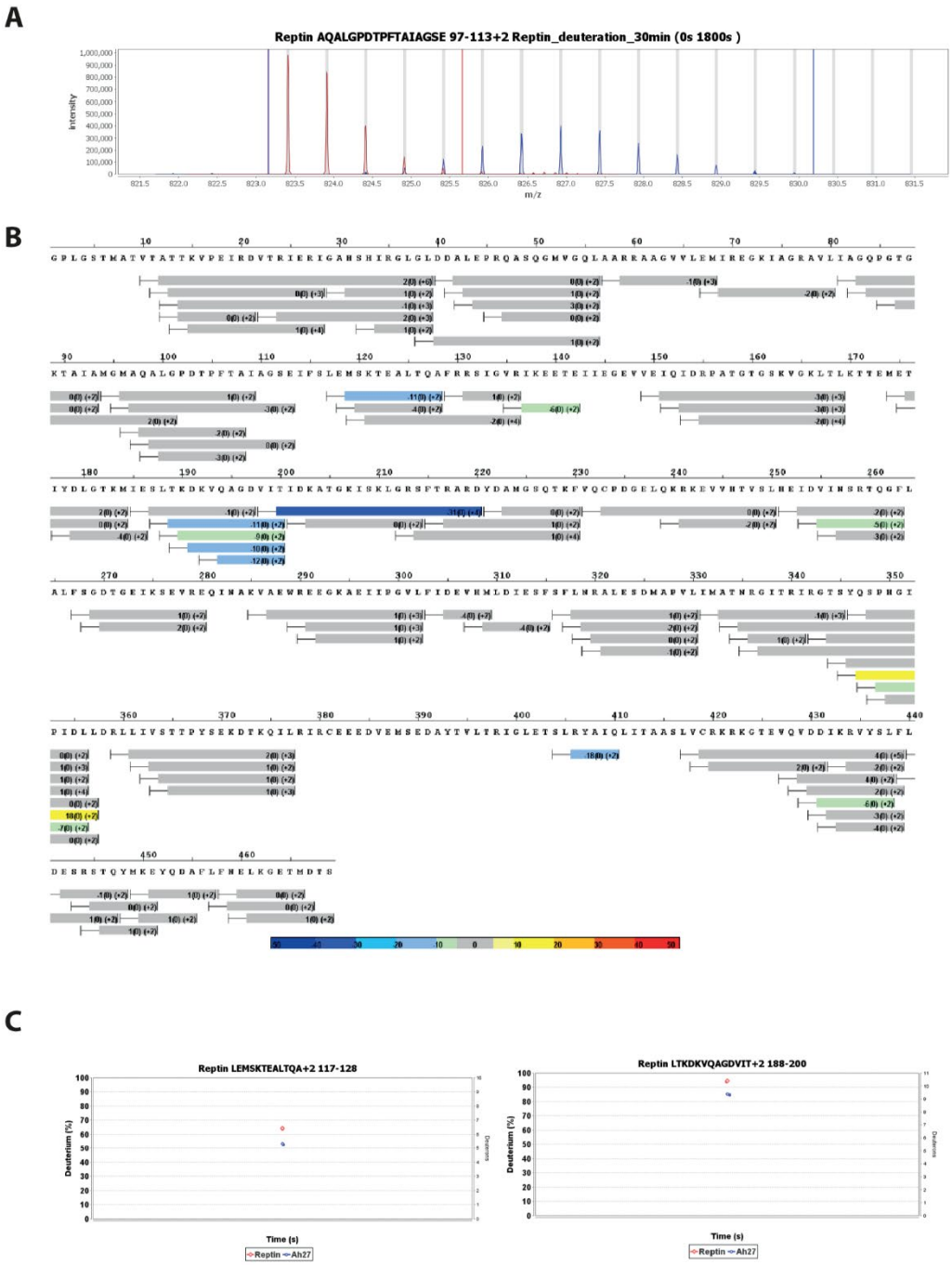


Figure 6.4 Deuteration levels on specific peptides of Reptin
A) Mass spectra for the peptide AQALGPDTPFTAIAGSE showing undeuterated peptide in red with the deuterated peptide (30-minute incubation) in blue. B) Changes in the levels of deuteration at 30 minutes comparing apo Reptin to Reptin incubated with Liddean for 1 hour prior to incubation with deuterated water. C) Changes in the levels of deuteration at 30 minutes comparing apo Reptin to Reptin of specific peptides.

the externalisation of a peptide into the aqueous environment. As very few changes are observed this is likely due to a conformational change occurring on ligand binding. Two peptides can be seen in figure 6.4C showing the percentage of deuteration observed in the same peptide in the absence (red) and presence (blue) of Liddean, both of these peptides show a reduction in deuteration levels showing that they are more internalised in the protein in the presence of Liddean.

These results of the peptides showing the greatest changes in the level of deuteration on the addition of Liddean to Reptin were then modelled onto 3UK6 (figure 6.5A). Here we can see that the dimer interface has a reduction in the levels on deuteration, indicating that Reptin is forming a higher-order oligomer. We can also see that there is a reduction in deuteration levels on one side of the ATP/ADP binding pocket while the other side appears to have an increase in deuteration. This indicates that Liddean can be seen to be binding into the ATP/ADP pocket, blocking deuterium exchange. When looking specifically at the pocket (in space fill, figure 6.5B) we can see, with the help of the placement of the ADP molecule that deep in the pocket there is a reduction in deuteration, however, the increase in deuteration can be seen along the 'tail' of ADP. When comparing this to the modelling of how Compound 1 would bind into the pocket (Chapter 3, figure 3.1D) we can see that this correlates with the predicted orientation of the molecule.

6.3.1 **Finding the 'CLICK' motif**

Further mapping of the increases in deuteration was performed on the amino acid and crystal structure of Reptin, these can be seen in figure 6.6. The changes visualised here represent a 60-second exposure to deuterium. Several peptides can be seen in yellow showing around a 2-fold increase in the levels of deuteration with the peptide showing the greatest increase (10-fold) shown in red (panels A and C).

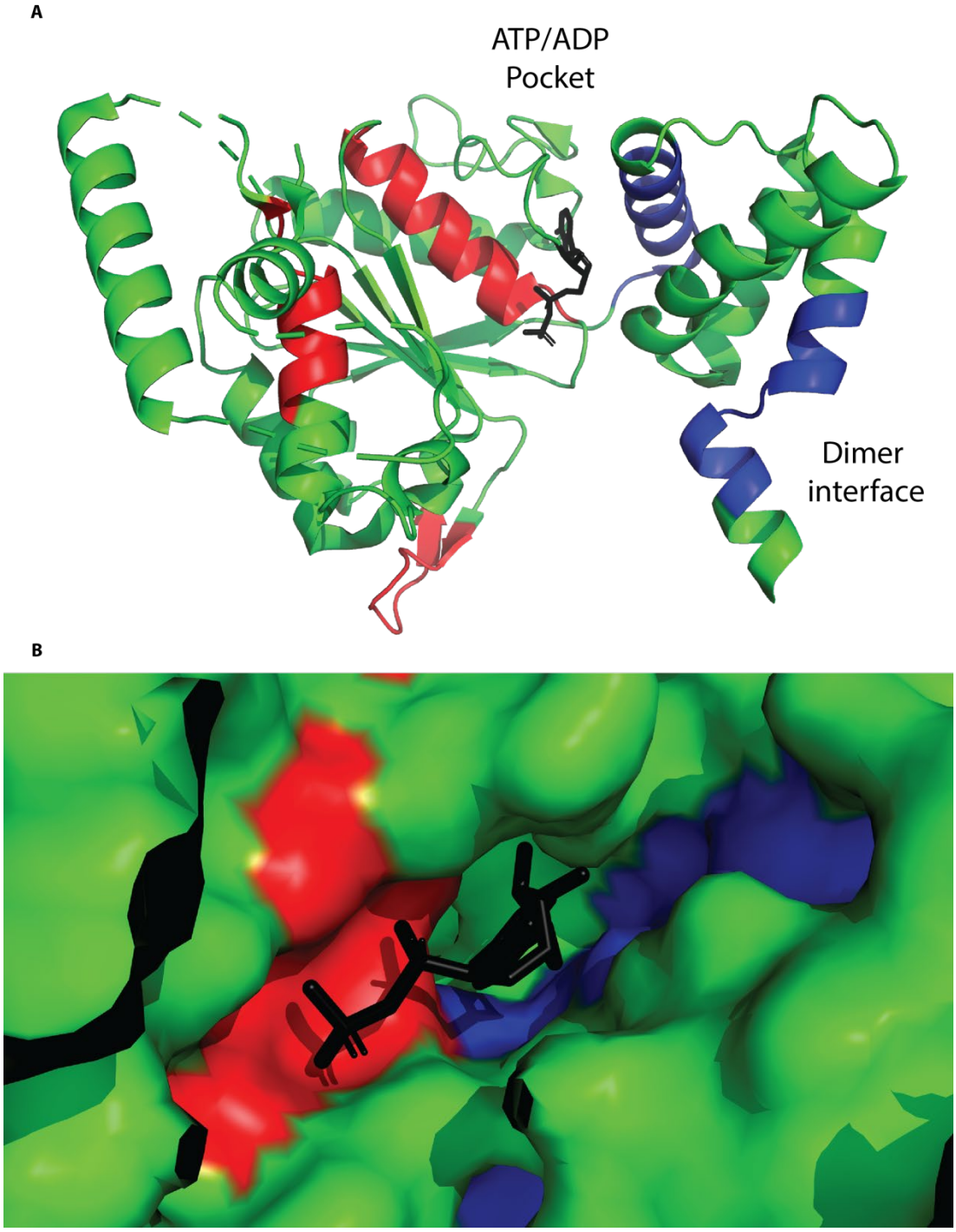


Figure 6.5 Mapping changes in deuteration due to Liddean binding.
The Reptin protein (PDB 3UK6 (Petukhov et al., 2012)) is displayed as a cartoon in green. ADP is black (sticks). Increases in deuteration on Liddean binding are red and decreases are in blue. A) Reptin monomer with dimer interface and ATP/ADP pocket highlighted. B) Surface rendering of 3UK6. Zoom of ATP/ADP pocket. (Healy et al., 2015).

A

```
MATVTATTKV PEIRDVTRIE RIGAHSHIRG LGLDDALEPR QASQGMVGQL     50
AARRAAGVVL EMIREGKIAG RAVLIAGQPG TGKTAIAMGM AQALGPDTPF    100
TAIAGSEIFS LEMSKTEALT QAFRRSIGVR IKEETEIIEG EVVEIQIDRP    150
ATGTGSKVGK LTLKTTEMET IYDLGTKMIE SLTKDKVQAG DVITIDKATG    200
KISKLGRSFT RARDYDAMGS QTKFVQCPDG ELQKRKEVVH TVSLHEIDVI    250
NSRTQGFLAL FSGDTGEIKS EVREQINAKV AEWREEGKAE IIPGVLFIDE    300
VHMLDIESFS FLNRALESDM APVLIMATNR GITRIRGTSY QSPHGIPIDL    350
LDRLLIVSTT PYSEKDTKQI LRIRCEEEDV EMSEDAYTVL TRIGLETSLR    400
YAIQLITAAS LVCRKRKGTE VQVDDIKRVY SLFLDESRST QYMKEYQDAF    450
LFNELKGETM DTS                                            463

Around 2-fold increase
Around 10-fold increase
```

B

Fold change in deuteration of ligand bound Reptin compared to apo Reptin at the same timepoint

Peptide	60 secs	300 secs
85-101	3.338687	2.506236
94-113	2.406045	1.9646
174-184	2.368801	2.560776
285-302	2.362673	2.540412
316-330	2.100891	2.125895
331-357	2.667112	2.465783
333-345	11.47516	5.172925

C

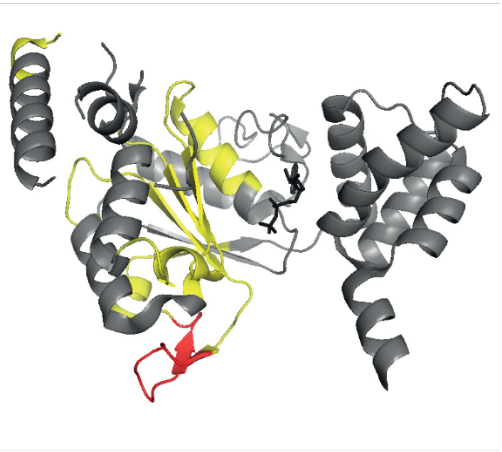


Figure 6.6 Mapping the increase in deuteration on Reptin
A) Amino acid sequence of Reptin showing greatest increases in levels of deuteration at 2-fold (yellow) and 10-fold (red) after 60 seconds of exposed to deuterated water. B) Fold changes deuteration of ligand bound Reptin compared to apo Reptin at the same time points observed for several peptides. C) Changes in A and B mapped onto the 3D structure (3UK6-A (Petukhov et al., 2012)) of Reptin monomer.

The table shown in figure 6.6B shows how some of these changes are dynamic, peptide 333-345, whereas the others are more stable. Even though the change seen in deuteration levels of peptide 333-345 are dynamic, at 300 seconds the levels are still 5-fold higher compared to around 2-fold with the other peptides. On the 3D mapping of the observed changes (figure 6.6C), it is clear that the part of the protein showing the greatest changes in deuteration levels could be interesting due to its placement on the edge of the Reptin structure.

As mentioned above, the peptide which was seen to have the greatest change in deuteration on the addition of Liddean, peptide 333-345. Due to the location of the peptide on the monomer, it was decided to look at its location in the full hexameric structure and be seen in figure 6.7. Here you can see this peptide marked in red or blue with the rest of the same monomer coloured paler on the 3D structure (3UK6). The peptide on monomer appears to lock into the same peptide adjacent monomer. In both the ribbon (figure 6.7A) and space fill (figure 6.7B) models the adjacent nature of these peptides is clear. The peptides appear to ’click’ into the next subunit, as seen in panel C, thus coining the term ’CLICK motif’ for this small peptide 333-345 which can be seen in red using stick modelling with the rest of the monomer observed using cartoon in beige, the adjacent monomer can be seen in blue in space fill mode highlighting the CLICK motif in a darker blue with the rest of the monomer in pale blue. On closer inspection it is possible to deduce that it is the tyrosine found at amino acid 340 which fits exactly into the pocket in the next monomer formed by the same peptide, panel C has this residue highlighted in yellow. The peptide 333-345 itself can be seen in panel D with an overlay of Reptin (red) and the corresponding Pontin (blue) peptide is found in panel E. What is most evident is that the tyrosine observed in the Reptin peptide (TRIRGTSQYSPHG) is absent from the Pontin peptide (CVIRGTEDITSPGH) and the central amino acids on

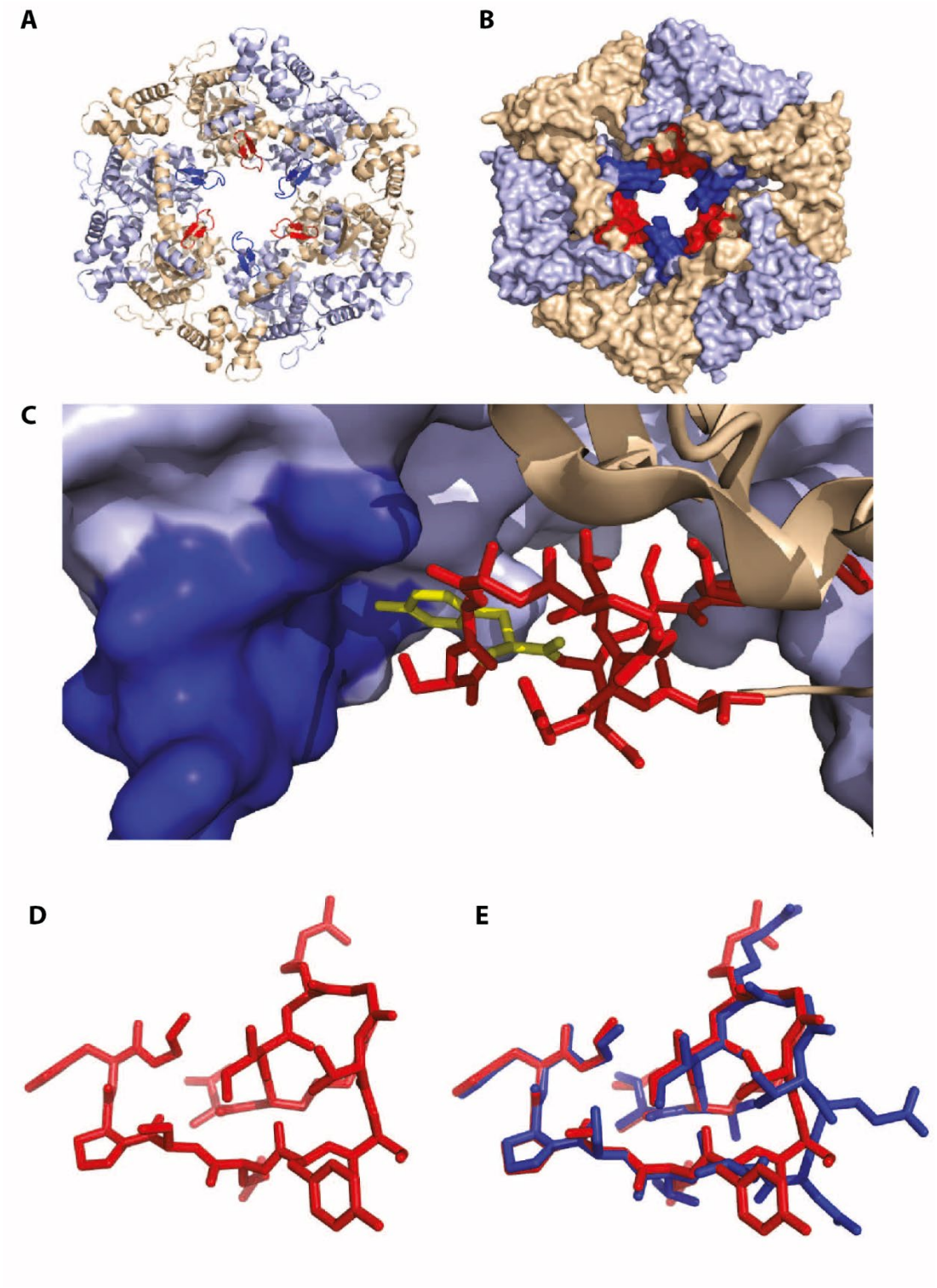


Figure 6.7 Modelling the CLICK motif

A) Cartoon of 3UK6 (Petukhov et al., 2012) hexamer with the CLICK motif highlighted in darker tones. B) Space fill of 3UK6 hexamer with the CLICK motif highlighted in darker tones. C) Zoom of CLICK motif seen in red using stick modelling with the rest of the monomer observed using cartoon in beige, adjacent monomer can be seen in blue in space fill, Tyrosine 340 residue highlighted in yellow. D) The peptide 333-345 of Reptin in stick mode. E) Overlay of the peptide 333-345 Reptin (red) and the corresponding Pontin (blue) peptide. (Remnant et al., 2019)

the Pontin peptide is very different although both contain a single hydrophobic residue. This could mean that the Reptin-Pontin heterododecamer likely forms in a similar, but not identical, way to the Reptin homohexamer.

Further investigation and modelling allowed for a greater understanding of the make-up of the click motif. The CLICK motif is localised within the ‘ring’ of the hexamer (figure 6.7A/B) and on an exposed surface of the monomer (figure 6.8A). The motif on one monomer appears to click into the adjacent monomer (figures 6.7B and zoom 6.8B). The peptide 333-345, TRIRGTSYQSPHG, can be found in figure 6.8C with several residues highlighted in a different colour; tyrosine (orange) in the pocket of histidine (blue), methionine (pink), arginine (red) and threonine-arginine-isoleucine-arginine (green).

Due to this interesting peptide, it was decided to create a point mutation in the wildtype protein by site-directed mutagenesis at the site of the tyrosine 340, mutating it to an alanine, this was simulated by modelling (figure 6.8D) showing the reduction in contacts between the two adjacent monomers. Primers were designed to bind to the pDEST15-Reptin plasmid to introduce the point mutation. After PCR with these primers, the product was digested with Dpn1 and then transformed into DH5α cells for purification. Sequencing was performed by Source bioscience to confirm the presence of the required mutation. Alongside this site-directed mutagenesis to have the D299N mutant in the same vector was also performed. D299N is a well characterised, catalytically dead form of Reptin that is known to form stable hexamers. The new mutant plasmids were transformed into BL21-AI cells and expressed as per the standard protocol. Purified protein was obtained for both mutants to allow for ongoing experiments. The characterisation of this new

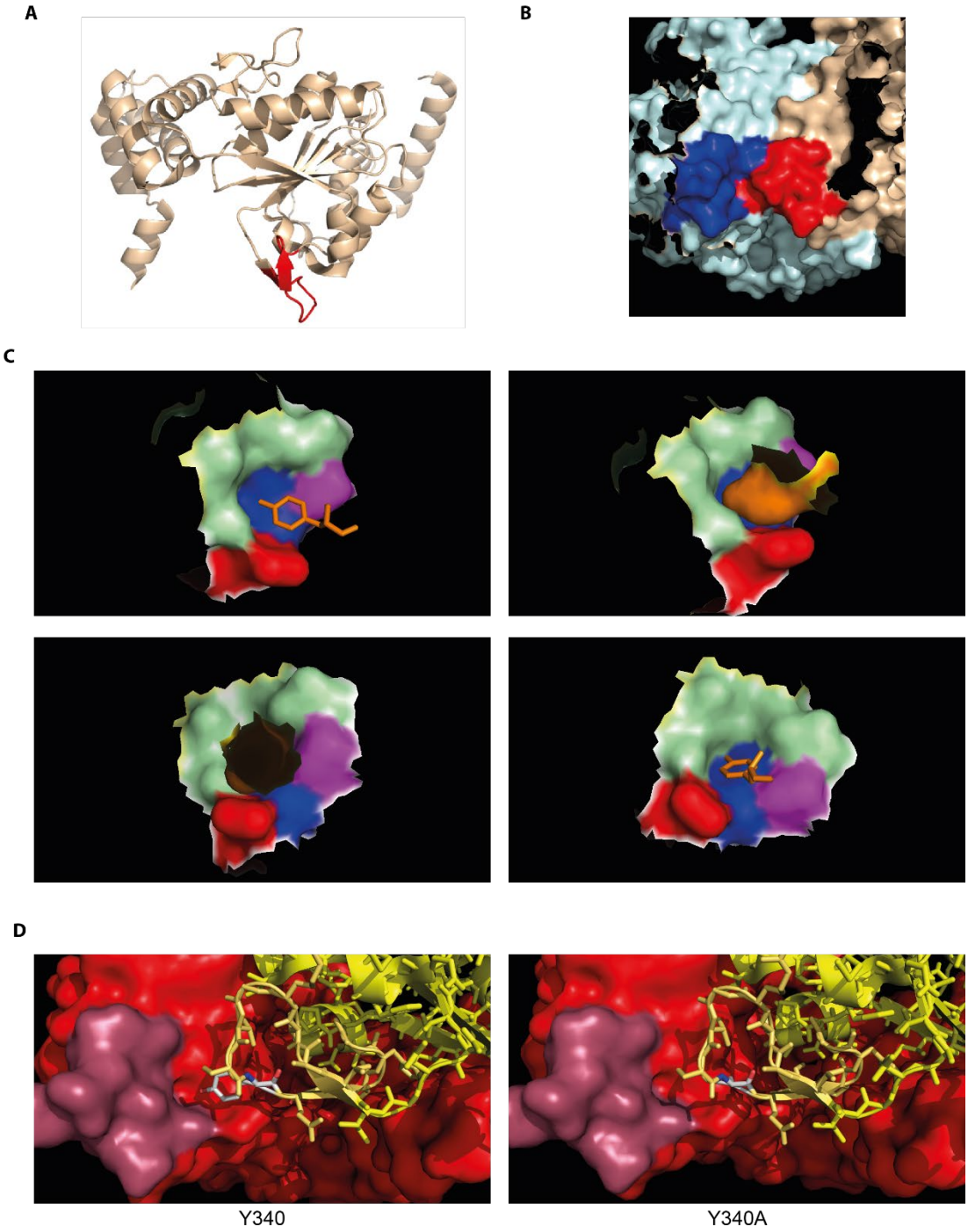


Figure 6.8 Make-up of the CLICK motif

A) Monomer of Reptin (3UK6 (Petukhov et al., 2012)) with the CLICK motif, peptide 333-345, highlighted in red. B) Space fill of 2 adjacent monomers of Reptin with the CLICK motif highlighted in darker tones. C) Tyrosine (orange) in pocket of histidine (blue), methionine (pink), arginine (red) and threonine-arginine-isoleucine-arginine (green). D) Tyrosine 340 in stick mode mutated to alanine. (Remnant et al., 2019)

protein will be done along with site the wildtype protein and the well characterised D299N mutant.

6.4 Separating oligomeric forms of Reptin

Due to now having three analogues of Reptin, the wildtype, the well-characterised D299N which results in a reduction in ATP hydrolysis and the new Y340A mutant, a wide variety of assays was needed to help characterise what these mutations might induce compared to with wildtype.

6.4.1 Gel filtration

Gel filtration was performed to see if we could observe changes in the oligomeric state of these proteins. A column of Sephadex was used and the proteins were added to the top of the column. Over time fractions were taken from the column and then analysed for the presence of protein by ELISA. This was performed in the presence and absence of 1mM ADP (large excess) to try and induce different oligomeric forms of Reptin. The results of these experiments can be found in figure 6.9 Panels A) wildtype, B) Y340A and C) D299N. The quicker a protein elutes on the column the larger it is as the smaller proteins take a long journey through the column. Molecular weight markers eluted as follows 440kDa in 11ml fraction, 158kDa in 14.5ml fraction, 43kDa in 14ml fraction and 13kDa in 18ml fraction. In the wildtype protein we can see two oligomeric forms of the protein with the addition of ADP, but only a single peak with the protein only, due to its elution size we can conclude that the peak seen in both is the monomeric peak. Comparing this to the Y340A and D299N mutants we can see that it would appear that the Y340A mutant is a single monomeric species, comparatively the D299N mutant appears to be able to form higher-order oligomers without the need for the ligand to be present.

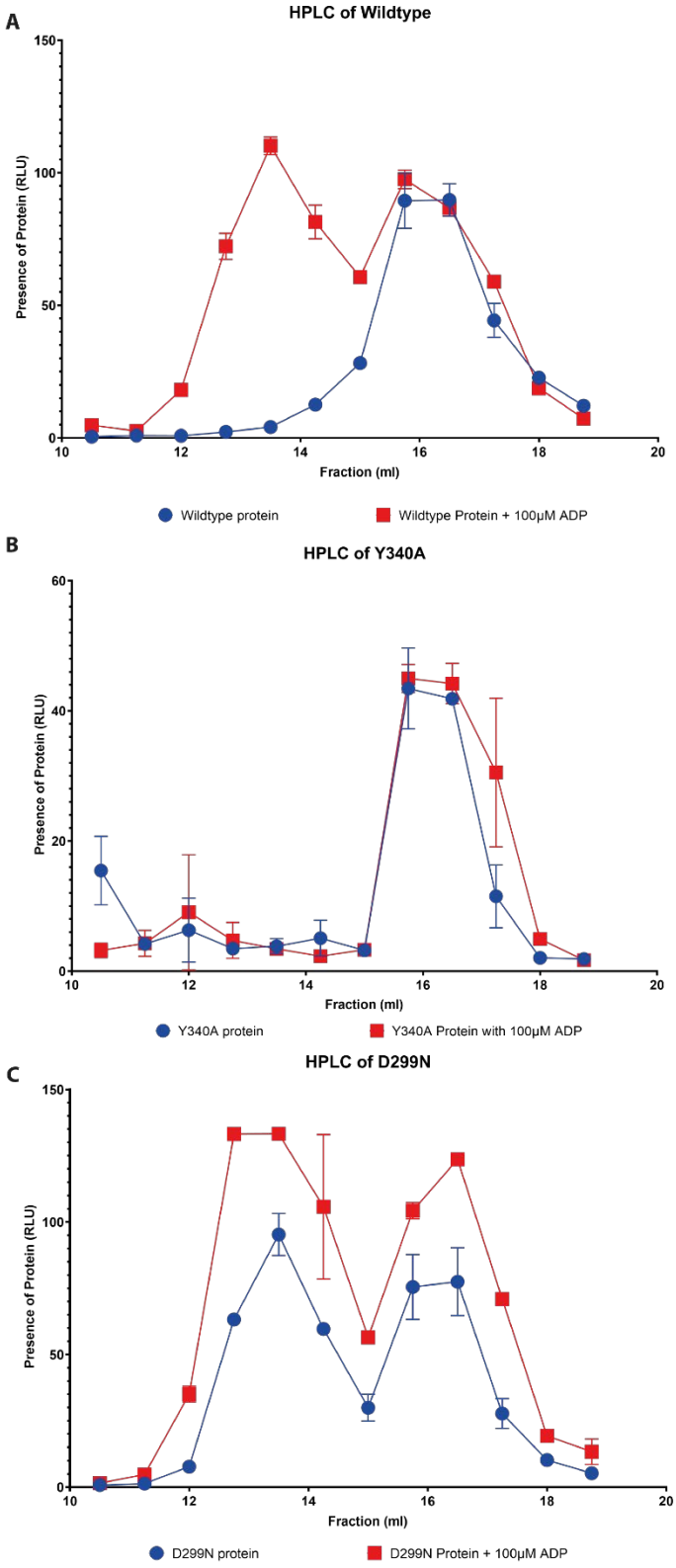


Figure 6.9 Gel filtration on Reptin in absence and presence of 100µM ADP. *Molecular weight markers eluted as follows 440kDa 11ml, 158kDa 14.5ml, 43kDa 14ml, 13kDa 18ml. A) Wildtype Reptin B) Reptin Y340A mutant C) Reptin D299N mutant. (Remnant et al., 2019)*

If this is the case that the Y340A CLICK motif mutant cannot form oligomers, then this gives us a much greater understanding of the importance of the CLICK motif to the structure of the protein.

6.5 Measuring changes in the activity of Reptin with the addition of point mutations

The ATP binding and ATPase activity of Reptin alone is chronicled as being very low. It is important to see if the changed in the Y340A mutant result in changes in these abilities. The increased affinity of D299N with ATP has been previously seen so this allows this protein to be used as a positive control for ATP binding compared to the wildtype protein.

As it is known that many AAA+ proteins only exhibit their ATPase activity when they are bound in complex, Pontin was now required for investigation. Using the Pontin sequence from the pCDNA-3xFLAG-Pontin (a kind gift from the Artandi lab (Venteicher et al., 2008)), Pontin was added to pDEST15 in the same manner as Reptin. The expression of Pontin was carried out in the same manner as Reptin this time with induction occurring at 30^{o}C for a greater yield of Pontin.

6.5.1 Measuring the binding affinity of Reptin and its variants for ATP

To perform this assay, Reptin, wildtype or mutant, is incubated with ATP-gamma32P in solution. The solutions are then filtered through a disc to capture the protein. If the ATP is binding to the protein it will be captured with the protein on the disc. These discs are then placed on a phosphor screen to expose the amount of radioactivity in each sample. The phosphor screen is then imaged using a Phosphoimager. This creates a digital image, similar to a western blot, which can be used for quantitation.

We can see in figure 6.10A can see that wildtype Reptin has a very low affinity of ATP, not much higher than background and the D299N has a much higher affinity, as expected. The Y340A mutant can be seen to have an affinity similar to that of the wild-type protein.

6.5.2 **Measuring the ATPase activity of Reptin and its variants**

The enzymatic ability of Reptin alone is known to be very low, it is thought that the enzymatic activity of Reptin and Pontin can only be observed when they are in the complex. Using a colourimetric assay for the amount of free phosphate we can examine the ability of the protein to break down ATP into ADP and phosphate. Here the protein was incubated for an hour before the experiment to ensure protein complexes had time to form, this was especially important for the Reptin/Y340A and Pontin.

We can see in figure 6.10B that the amount of free phosphate produced by Reptin wildtype and Pontin in the complex is 4-fold higher than Reptin or Pontin alone. The combination of Pontin with the Y340A mutant does not seem to have an increase or decrease in the amount of free phosphate available compared to the proteins alone. Figure 6.10C shows us that changing the amount of free ATP used in the experiment does not affect the results seen, the amount of free phosphate produced by the complexes increases with the amount of ATP input but the Y340A and Pontin proteins together do not create an amount of free phosphate similar to that of the wildtype protein with Pontin.

6.5.3 **Measuring Reptin and mutant binding affinity for AGR2**

A protein-protein interaction assay was carried out by coating AGR2 onto a polystyrene plate (as seen in Chapter 3). Here we titrated increasing amounts of wildtype and mutant Reptin to observe the changed in binding affinity. In figure 6.11

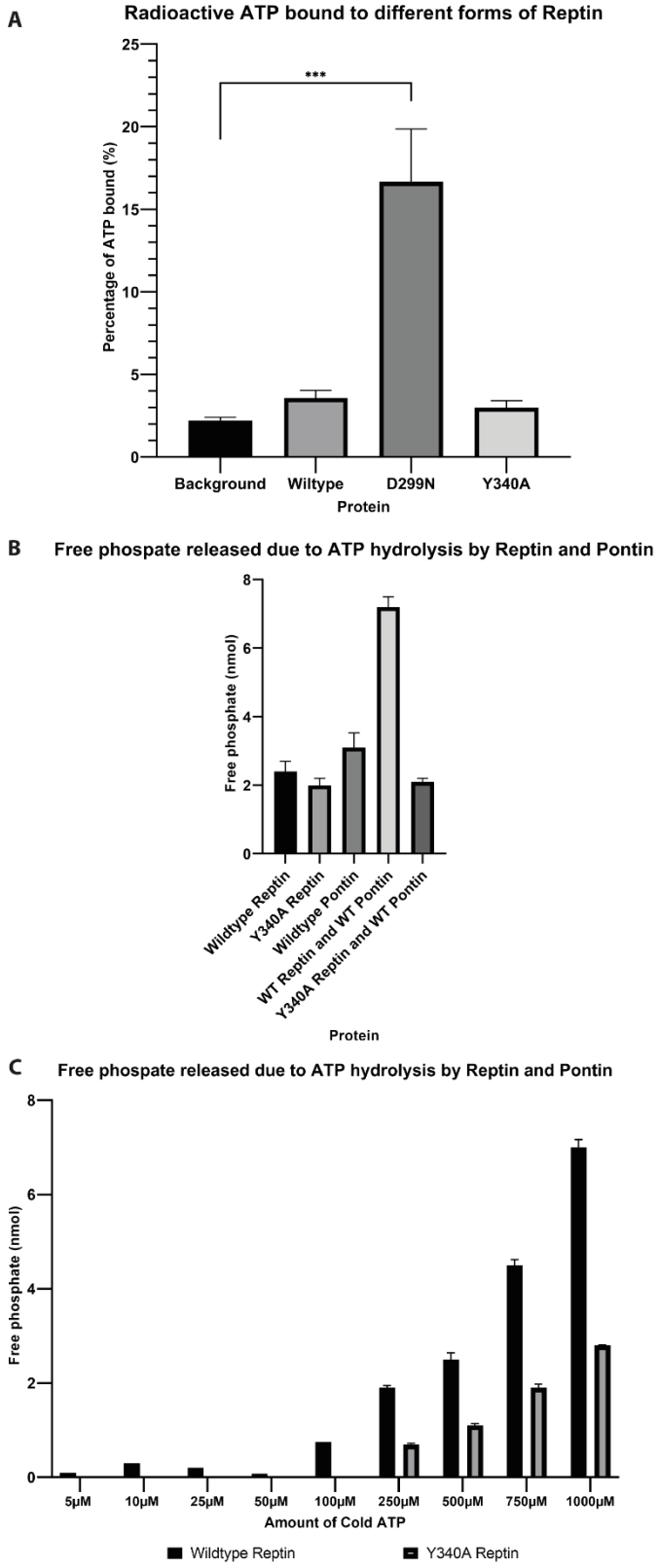


Figure 6.10 ATP binding affinity of Reptin.
ATP-gamma32P was incubated with different forms of Reptin before being captured on a nitrocellulose filter. Quantification occurred using Phosphoimager.

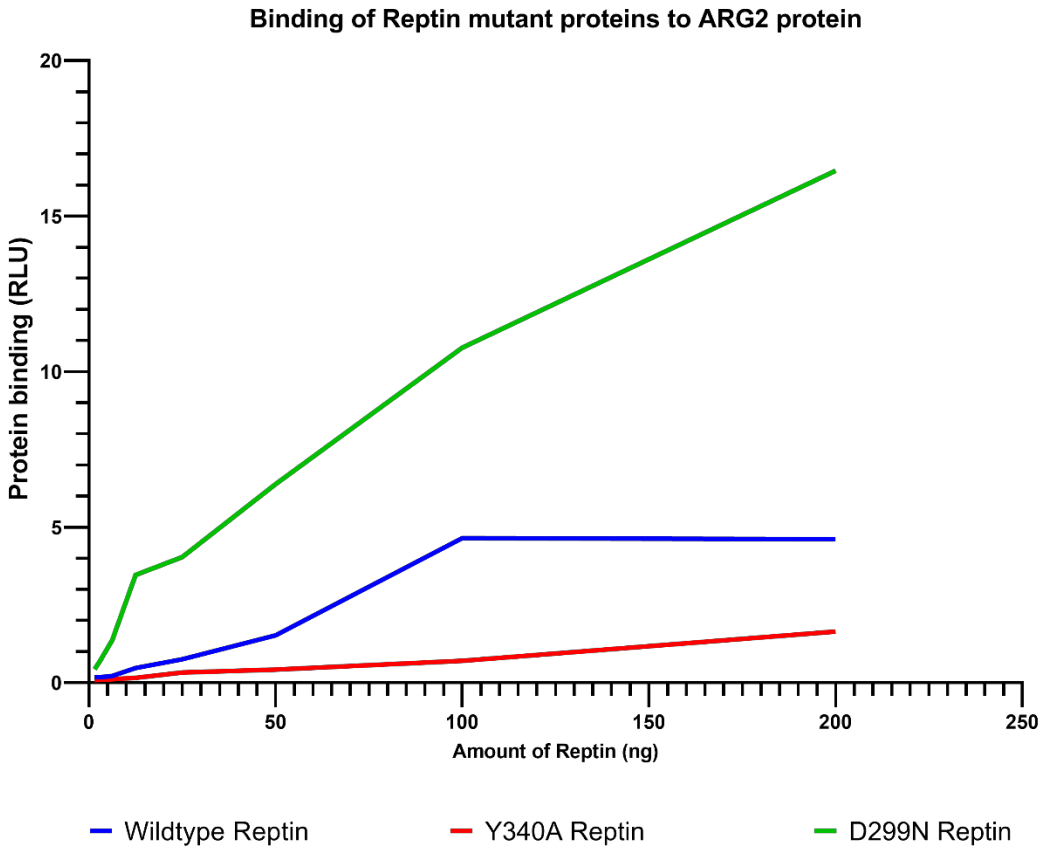


Figure 6.11 Measuring Reptin and mutant binding affinity for AGR2.
Wild-type Reptin and the indicated mutants were analysed in AGR2 protein binding assays. AGR2 (1µg) was pre-coated onto ELISA wells. Reptin was titrated, and anti-Reptin and HRP-conjugated secondary antibodies were used to measure the extent of Reptin binding. (Remnant et al., 2019).

it is evident here that the Y340A mutation increases the affinity of Reptin for AGR2 whereas the D299N mutant acts similarly to the wildtype protein.

6.6 Discussion

This chapter has shown how a combinatorial approach can lead to a much better understanding of the molecular conformation and oligomerisation of the protein of interest. Hydrogen deuterium exchange, in combination with the binding of the novel ligand Liddean, has led to the discovery of a novel site on Reptin that has been further investigated. The Reptin peptide 333-345 TRIRGTSYQSPHG showed the most dramatic increase in deuteration, 10-fold at 60 seconds, compared to all other peptides in the protein. This means that this peptide is likely to be more enclosed in the monomeric form of Reptin. Apparently, during oligomerisation, this peptide was released from the monomeric conformation to aid in the formation of a stable oligomer.

Modelling of this peptide, TRIRGTSYQSPHG, revealed that it might help form a pocket for the same peptide in the adjacent monomer to bind into. Thus, leading to the 'CLICK' motif, designated as this due to the clicking together of the peptide into the same peptide on the adjacent monomer. The tyrosine in the peptide (TRIRGTS**Y**QSPHG) appeared to be crucial for the formation of the hexameric Reptin ring. Further investigations allowed for confirmation me to confirm the importance of this amino acid in the oligomeric structure of Reptin.

In Chapter 3, I showed that a specific self-peptide, peptide 28, of Reptin have a very high affinity for Reptin protein binding. The sequence of peptide 28 is GITRIRGTSYQSPHG. If we compare this to the CLICK motif, TRIRGTSYQSPHG, we can see that they are near identical, with peptide 28 containing two additional amino acids at its N-terminus. As this peptide was seen to have such a high level of binding to Reptin protein (self-binding). This observation further helps to confirm the importance of this motif. This peptide also shows the greatest reduction in binding when Reptin has previously been incubated with ligands (ADP and Liddean)

suggesting it may be important for the oligomerisation on the protein (Chapter 3, figure 3.5).

I have found that different oligomeric forms of Reptin can be visualised by gel filtration. The wildtype protein can form higher-order oligomers with the addition of the endogenous ligand, however, the Y340A mutant is not able to form a larger stable oligomer. The D299N mutant was able to form a higher-order oligomer without the need for ligand binding which was expected, based on previously published data (Maslon et al., 2010).

Reptin is a member of the AAA+ superfamily (Rosenbaum et al., 2013). These are known for their oligomerisation on ATP binding and ATPase ability (Hanson and Whiteheart, 2005)]. I have tested the wildtype and mutant proteins for both. The experiment looking at the binding affinity of the Reptin mutants for ATP justified the hypothesis that it is the oligomeric form that is required for higher binding to ATP. Thus, the D299N mutant has a higher affinity for ATP compared to WT and Y340A Reptin (figure 6.9). The ATPase experiment also showed that in the presence of equimolar amounts of Pontin, wildtype Reptin is able to convert ATP into ADP + Pi. This was not the case for the Y340A mutant and is likely due to its inability to form an oligomer with Pontin, although this was not tested.

My data reveal that residue Y340 of Reptin is required for oligomerisation to occur. Without oligomerisation, Reptin is unable to bind and hydrolyse ATP. We have previously seen that the binding partners of Reptin, AGR2 and p53, have different binding profiles with wildtype Reptin in chapter 3. The binding of Y340A Reptin to AGR2 is much higher than the binding of either wildtype or D299N Reptin. As my data suggested that Reptin Y340A is the monomeric form of Reptin, this also agrees with the reduction in binding between wildtype Reptin and AGR2 seen in the

presence of ligand. This possibility could be further investigated in the future using the monomeric Reptin specific ScFv discovered in Chapter 4. This antibody would allow investigation of monomeric Reptin in cells and its localisation and colocalization with AGR2 and p53. This could be achieved a combination of imaging techniques including immunofluorescence (Odell and Cook, 2013), FRET/FLIM (Wallrabe and Periasamy, 2005) and proximity ligation assays (Fredricksson et al., 2002).

As previously mentioned in Chapter 4, it is thought that monomeric Reptin may be oncogenic. I have previously shown in Chapter 3 that the binding between Reptin and AGR2 is decreased in the presence of ligands. Reptin binding is thought to activate AGR2, which in turn inhibits p53. In contrast, the binding between Reptin and p53 is increased in the presence of ligands and is thought to activate its tumour suppressor abilities. The identification of a specific peptide in Reptin that could be responsible for its transition between monomer and oligomer be relevant in a clinical setting due to the high incidence of overexpression of Reptin in cancer patients, in particular hepatocellular carcinoma (Rousseau et al., 2007, Iizuka et al., 2006), colon cancer (Graudens et al., 2006), melanoma (Talantov et al., 2005), bladder cancer (Sanchez-Carbayo et al., 2006) and prostate cancer (Kim et al., 2005, Kim et al., 2006).

One possible speculative explanation for the increase in the levels of monomeric Reptin could be that if p53 binding helps to stabilise the Reptin oligomer in vivo when the increase in the overall expression of Reptin in the cell is not accompanied by a similar overexpression of p53, there might be a decrease in oligomer stability resulting in a greater proportion of monomeric Reptin. This increased monomeric Reptin could then preferentially bind to AGR2, activating it, resulting in increased inhibition of the remaining pool of p53 in the cell.

To see if this has any clinical relevance, it would be possible to investigate specific mutations seen in cancer patients and compare their localisation in both the Reptin sequence and the 3D structure of Reptin. Any mutations found at the inter-subunit interface or in the rest of the CLICK motif would be of particular importance.

Here, Reptin is used to develop new approaches to dissect oligomerisation driven by ligand binding. This method utilises hydrogen-deuterium exchange to highlight sites on the external faces of a protein which are exposed or enclosed on small molecule binding. A single binding motif for Reptin was identified using the HDX experiments, identifying an important stabilising side chain the Y340 which is responsible for holding the homo-hexamer together. Combining self-peptide motif screens and HDX data allowed definition of an inter-subunit bridge formed between the same peptide in all monomers. A study looking at both Reptin and Pontin found that their oligomerisation and activity can be controlled through binding to Histone H3 N-terminal tails (Queval et al., 2014). This highlights the importance of binding peptides in the control of the oligomeric state.

Chapter 7 Discussion

7.1 Discovery and testing of Liddean

Discovery and testing of Liddean In Chapter 3 my work involved the first in vitro testing of compound output from the in-silico screening software LIDAEUS (Taylor et al., 2008). The software docked a virtual library of 4.4 million conformers. 49,971 hits identified as binding to the Walker A site of Reptin were re-docked using Autodock software (Trott and Olson, 2010). The hits were prioritized using a 'rank-by-rank' protocol and 30 were selected to be tested in protein-protein binding assays to assess the effectiveness of the *in silico* approach. From this 30, a single compound, Compound 1 (Chapter 3, figure 3.1C and D) was identified as having the greatest inhibitory effect on the binding of Reptin to known protein interactor, AGR2 (Maslon et al., 2010). Compound 1 was subjected to a structure-activity-relationship study (Chapter 3, figure 3.1E). The panel of Compound 1 derivatives was then subjected to the same assays resulting in the selection of Liddean for further studies. The binding of Liddean to the Walker A site of Reptin was readily detailed using hydrogen-deuterium exchange mass spectrometry (Chapter 6, figure 6.5A and B).

This approach allowed me to confirm the effectiveness of the LIDEAUS software to identify small molecule ATP mimetics, which bind to the Walker A site of Reptin. To extend this study, a larger number of proteins could be tested for changes in protein-protein interactions induced by small-molecule binding. Candidates would include Pontin, the homolog of Reptin, as well as other known binders of Reptin in these assays.

Similar approaches using in-silico screening and in vitro assays could be used in the future to find small molecule ATP mimetics, able to bind Walker A sites of other

AAA+ proteins for which crystal structures are available. Modelling of other Walker A sites and comparing them with the same conformer library could allow for a greater understanding of the specificity of these small molecules, perhaps allowing for some to be discounted if they were predicted to bind the Walker A sites of several proteins.

In addition to this, alternative assays could be used to look at changes in Reptin functional dynamics. For example, Reptin and Pontin are found together in many chromatin remodelling complexes. We observed in the proximity ligation assay that

Reptin-Pontin foci shuttle out of the nucleus into the cytoplasm (Chapter 4, figure 4.4C). The effects on Reptin-Pontin activity could be investigated further by looking specifically at chromatin remodelling (Flaus and Owen-Hughes, 2001). Well documented assays that could be used include ATP-dependant nucleosome remodelling (Chen et al., 2014), mono/di-nucleosome binding (Owen-Hughes et al., 1999, Chen et al., 2014), DNA and nucleosome dependant ATPase assays (Chen et al., 2014). These assays could be performed in the presence and absence of Liddean, using multiple Reptin-containing chromatin remodelling complexes. If changes in chromatin remodelling were observed, the next step would be to look for changes in gene expression in cells exposed to Liddean for many time points, this could be studied using RNAseq. This is computationally complex but could give rise to candidate genes (Finotello and Di Camillo, 2015, Rapaport et al., 2013). Those candidate genes could then be used to see if they undergo repositioning within the nucleus which can also be related to their expression level (Egecioglu and Brickner, 2011).

For work in vivo, the specificity of Liddean is of great importance. As much of the work here was performed in vitro with bacterially expressed proteins, we are not

able to comment on the off-target effects caused by Liddean in vivo. However, some off-target effects were likely observed in the experiment looking at changes in the electrical impedance of cells on the addition of Liddean (Chapter 3, figure 3.7) which were thought to be caused by changes in cell morphology and adherence capability.

One method to test the specificity of Liddean could be to make a derivative that can be chemically bound to beads, incubating the Liddean beads with cell lysates and performing mass spectrometry analysis on the bound proteins. This could yield a more specific target pool of possible interacting proteins which could be used to narrow down the list of off-target effects. To see if the changes in cell morphology and adherence a result of changes in the protein-protein interactions of Reptin are merely induced by Liddean, Reptin knockdown could be performed followed by treatment with Liddean. To enable quick removal of Reptin, the auxin-inducible degron (AID) system could be utilised (Nishimura et al., 2009). Here, rapid and reversible degradation of target proteins is possible due to the introduction of the plant-specific auxin pathway into animal cells, resulting in polyubiquitination of the target protein followed by degradation by the proteasome. If the result of this experiment was similar to those seen in the cells with normal levels of Reptin, the resulting changes in morphology could be attributed to an off-target effect of this small molecule. To test if this off-target effect was caused by the binding of Liddean to another AAA+ protein, such as dynein (Vale, 2000), a library of AAA+ AID cell lines could be made. Such AID cell lines produced could also be used for many of the other experiments outlined above.

7.2 Use of ScFv to identify a Reptin self-binding site

In Chapter 4 my work involved the use of antibody phage display, to identify 5 single-chain variable-fragments (ScFv) that bind Reptin. The binding site for ScFv 4 is not visible on the surface of the Reptin homohexamer (Chapter 4, figure 4.10B),

whereas the sites for ScFv 25 and the multi-ScFv site are seen on the external surface (PDB 3UK6). This means that ScFv 4 is unlikely to be able to bind to oligomeric Reptin. In the assay looking specifically on the effect of the presence of Liddean on ScFv binding, the binding of ScFv 4 was lost (Chapter 4, figure 4.7). This was expected as we have seen from the gel filtration that Reptin bound to the ligand is an oligomer (Chapter 6, figure 6.9). This further argues that us that this ScFv is only able to bind to monomeric Reptin.

To investigate this further, we could use the mutant proteins expressed for the experiments in Chapter 6. If the Y340A mutant is monomeric, the ScFv 4 should be able to bind successfully in the presence and absence of ligand. In contrast, D299N should have very low, if any, binding of ScFv 4 due to its oligomeric conformation. Wildtype Reptin could also be used to test this using the addition of various ligands.

To help identify the binding site of ScFv 21, which binds Reptin but not the Reptin overlapping peptide library we could use the hydrogen-deuterium exchange to map the binding site, as we did for Liddean (Chapter 6, figure 6.5). This could confirm that ScFv 21 has a structural epitope rather than a linear peptide epitope. This technique could also be used to map the binding sites of the other ScFv, giving information on the oligomeric state of the proteins that are bound to the specific ScFv.

If it can be confirmed that ScFv 4 is specific for monomeric Reptin then this could be used in proximity ligation assays to ascertain the oligomeric state of Reptin when it is bound to each of its binding partners. If other of the ScFv were confirmed to bind specific oligomeric states of Reptin these could also be used in similar assays.

7.3 Use of peptide motif screen to identify Liddean dependent motifs

In Chapter 5, I used peptide phage display to look at changes in the binding landscape of Reptin on ligand binding. In the first screen, I identified specific binding motifs from the dataset, one of which turned out to be relevant to the interaction of Reptin with p53. On further investigation, the motif was found to be present in a peptide mapping to the tetramerization domain of p53. The p53 protein is only active in its tetrameric form, which binds to DNA with high affinity. The activation of p53 is closely linked to its conformation with a regulatory domain next to the tetramerization domain at its C-terminus. As Reptin is found in several chaperone complexes, such as R2TP, the interaction may help to ensure the correct conformation of p53 allowing it to carry out its role as a tumour suppressor. To investigate this further, investigations into this complex involving looking at the molecular ratio of the two proteins using techniques such as ion-mobility mass spectrometry (Kalenius et al., 2018), which gives information on the size and shape of a complex, and SEC-MALS (Tarazona and Saiz, 2003), which allows for estimation of the size of protein complexes. It would also be interesting to test the DNA binding ability of p53 in the presence of Reptin, Pontin, and the two proteins together. It would also be possible to locate the binding site on Reptin for p53. This could be achieved using the Reptin specific peptide library performing peptide-protein interaction assays and hydrogen-deuterium exchange using peptide 38 of p53 and Reptin.

As p53 is well characterised for its mutations in cancer, it would be of interest to investigate whether any of these mutations are found at the Reptin binding site and if, when mutated in bacterially expressed proteins, they abolish the binding between Reptin and p53. It would also be interesting to know if Reptin binding to mutant p53

could be stimulated by the addition of Liddean, forcing Reptin into its hexameric form which appears to have a higher binding affinity for p53.

The identification of several peptides that can bind to Reptin (Chapter 5, figure 5.6), and whose binding is affected by the addition of Liddean, could also lead to other lines of further experimentation. Initially, it would be of interest to ask whether those peptides which can be seen to have a high affinity for either Reptin, Reptin with Liddean or both occur elsewhere in the human proteome, possibility identifying novel protein-protein interactions of Reptin. It would also be interesting to see which of these peptides can bind to the mutant forms of Reptin Y340A and D299N, and how these levels of binding compared to wildtype Reptin. This could also be tested in the absence and presence of Liddean and other ligands. Mapping of these binding sites could be achieved with the hydrogen-deuterium exchange. If specific peptides are seen to bind to oligomeric Reptin, it would be very informative to determine if they bind across multiple monomers of Reptin helping to 'lock' Reptin into its oligomeric form.

7.4 Fine mapping of Liddean induced changes in Reptin

Hydrogen-deuterium exchange mass spectrometry (HDX-MS) was used to compare wildtype Reptin to Reptin bound to Liddean. As discussed earlier I was able to map Liddean binding into the ATPase site of Reptin. Comparison of the predicted binding of Compound 1 (Chapter 3, figure 3.1D) to the deuterium mapped ATPase site (Chapter 6, figure 6.5B) revealed that Liddean is bound in the predicted manner. This shows that HDX-MS can be used to effectively map binding sites on a protein of interest.

To analyse the localisation of deuteration on the Reptin monomer, the sites were mapped onto the PDB 3UK6 structure of Reptin (Chapter 6, figure 6.7A and B). A single peptide -333-345- showed much higher levels of deuteration than the others (Chapter 6, figure 6.8). On closer inspection of the contacts of this peptide, I noticed tyrosine 340 protruding into a pocked in the same motif on the adjacent monomer (Chapter 6, figure 6.7C). I also compared the corresponding peptide in Pontin (Chapter 6, figure 6.7E) and noted that this lacked the tyrosine. In its place is isoleucine, which forms similar contacts with the adjacent monomer. Modelling the conversion of tyrosine 340 to alanine (Chapter 6, figure 6.8D) showed that this mutation would likely result in the loss of several contacts between adjacent monomers. The bacterial expression vector for Reptin was mutated using site-directed mutagenesis to give mutant Y340A Reptin. At the same time, I also mutated D299N, a mutant of Reptin which was previously characterised as a 'loss-of-function' mutant unable to hydrolyse ATP.

Several biochemical assays of wildtype Reptin, Y340A and D299N revealed that Y340A is unlikely to be able to form a hexamer. The gel filtration experiment was carried out 4 times, each time revealing that wildtype Reptin could be visualised as two peaks in the presence of ligand but Y340A was only ever able to form a single peak. I also used sedimentation assays (performed in duplicate) but could not successfully resolve multiple peaks for any form of Reptin. This may be due to the high G force and vacuum used in this experiment, resulting in a collapse of the oligomeric form or it could be due to Reptin being separated from ligands required to form oligomers in the gradient. The sedimentation assay was carried out over 18 hours in comparison to only 50 minutes for the gel filtration experiment. These various factors could explain why it was possible to observe the different oligomeric forms in one experiment but not the other.

Functional assays were then carried out on Reptin wildtype and mutants revealing that the D299N mutant can bind ATP in the absence of Pontin but Y340A and wildtype Reptin are not. However, in the presence of Pontin, wildtype Reptin can hydrolyse ATP which the Y340A mutant is not. This is likely due to the inability of Y340A to form a heterohexamer with Pontin, again reflecting the importance of this residue.

The final experiment was to perform HDX-MS on the Y340A mutant and compare the levels of deuteration to wildtype Reptin in the absence and presence of ATP. Interestingly no changes of deuteration were detected in the 'CLICK' motif (-333345-) in the Reptin Y340A mutant when exposed to the ligand. The lack of changes on the deuteration levels of the 'CLICK' docking motif in Reptin Y340A protein is consistent with the model that loss of Tyrosine 340 results in the Reptin monomer as in the wildtype protein this peptide is seen to increase in deuteration after the addition of ligand.

It may be possible to test if the binding affinity for Liddean is different between the different Reptin mutants as they appear to be more stable in their confirmation that wildtype Reptin.

7.5 A new model for the Reptin binding landscape

Integrating the diverse data gained during this thesis it is possible to postulate a model for the binding of Reptin to its various binding partners in the presence of different ligands. I have shown that ligand binding stimulates oligomerisation of Reptin either alone as a homohexamer or in complex with Pontin as a heterododecamer.

In figure 7.1, monomeric Reptin can bind AGR2, there may also be binding partners for monomeric Pontin. When stimulated with ligand (ATP/ADP/Liddean), Reptin and

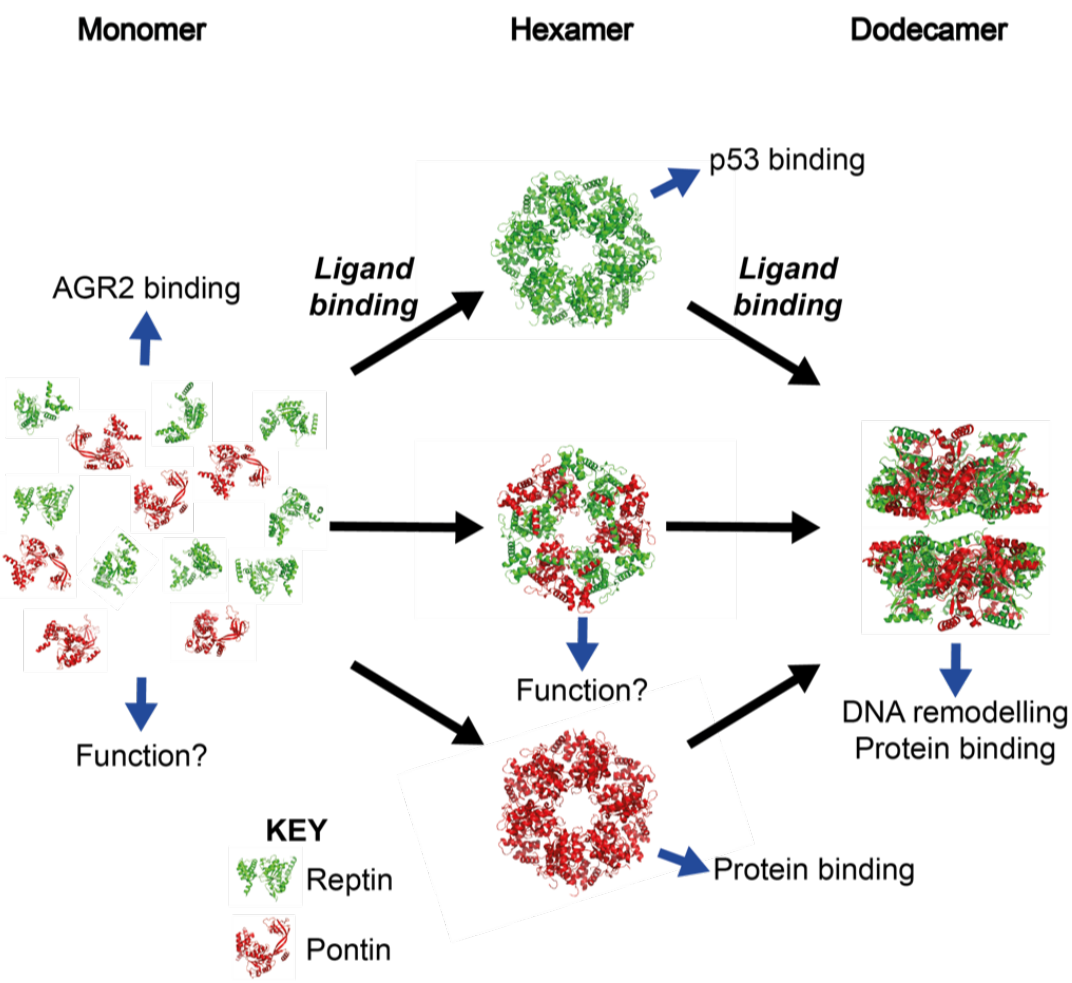


Figure 7.1 Model for Reptin binding to specific partners in its different oligomeric states
Reptin is in green and Pontin is in Red

Pontin can form hexameric structures alone in vitro and both have been solved by x-ray crystallography. It is not known whether the Reptin-Pontin complex can form a stable heterohexamer. These two, possible three, oligomeric forms of Reptin and Pontin could therefore have their binding partners independent of one another, suggested here by the binding of p53 to oligomeric Reptin. The heterododecamer formed by Reptin and Pontin is well characterised as having roles in DNA remodelling and is known to be part of many protein complexes. This is likely the most abundant oligomer for Reptin and Pontin. More investigation into the prevalence and interactions of the different oligomeric forms of these two proteins would lead to a much better understanding of Reptin and Pontin functions and in particular why they are found to be upregulated in various cancers. For example, Reptin is overexpressed in hepatocellular carcinoma (Rousseau et al., 2007, Iizuka et al., 2006), colon cancer (Graudens et al., 2006), melanoma (Talantov et al., 2005), bladder cancer (Sanchez-Carbayo et al., 2006) and prostate cancer (Kim et al., 2005, Kim et al., 2006). The techniques used here could likely also be applied to other AAA+ proteins.

7.6 Drugging the undruggable

The term undruggable is used to describe proteins that are currently unable to be targeted pharmacologically. Current developments in techniques mean progress is being made to target these proteins, many of which are oncogenes (Dang et al., 2017). The ability to target these ‘undruggable’ proteins would mean greater specificity for therapeutics as it is often these proteins which fall at crucial steps in signalling pathways, for example, proliferation and apoptosis (Hantschel et al., 2011). A large number of these proteins are transcription factors, such as MYC, MYB (Dang, 2012) and

nuclear-factor κB (Baud and Karin, 2009) and many cancers are found to have mutations in these proteins. A second group is the RAS family of proteins (Cox et al., 2014) who is found to be mutated in around 30% of cancers. Protein phosphorylation is another possible target for therapeutics, but this can be difficult due to the complex signalling pathways they are involved in and single kinases being able to target multiple proteins (Bain et al., 2007). The ability to selectively inhibit regulatory subunits could be important here as currently, many inhibitors are designed to block phosphorylation by the kinase. To target all the 'undruggable' proteins it is not possible to use the standard methods used for drug development.

The methodology adopted in this thesis allows observation of available small molecules and the ability to then develop them for specificity allowing them to be optimised for use with proteins which were not their original target. For example, despite many years of research, there have been no approved drugs against the protein KRAS. It was recently discovered that there is a shallow pocket of the surface of active and inactive KRAS but due to the shape it was considered undruggable (Quevedo et al., 2018, Cruz-Migoni et al., 2019). The use of LIDAEUS and peptide phage display techniques would allow for the creation of specific small molecules or peptides which can bind directly to this shallow pocket. A similar project was undertaken in the manner of our screen (Healy et al., 2015), where they developed an inhibitor of KRAS which blocks downstream signalling and has an antiproliferative affect (Kessler et al., 2019). In this study they observed the shape of the pocket and developed small molecules which could bind this surface. They

were able to use a ‘structure activity relationship’ study, as we did, to optimise the small molecule and used co-crystalisation to look at binding of the small molecule to KRAS. This is an alternative method to our HDX experiments giving high resolution data for the binding of the small molecule to KRAS (Kessler et al., 2019) which we were not able to obtain.

The ability to target proteins and protein-protein interactions which before were classed as ‘undruggable’ has been achieved with the development of *in silico* screening and with biophysical techniques. The ability to model binding pockets and interfaces means that it is now possible to develop molecules which are highly specific for the site on a particular protein which they have been designed to bind to. This allows sites which would not normally be classed as ‘binding pockets’ to be used to create changes in oligomerisation, conformation and protein-protein interactions.

Chapter 8 References

AITKEN, A. 1999. Protein Consensus Sequence Motifs. *Molecular Biotechnology,* 12**,** 241-253.

BAILEY, T. L., BODEN, M., BUSKE, F. A., FRITH, M., GRANT, C. E., CLEMENTI, L., REN, J., LI, W. W. & NOBLE, W. S. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res,* 37**,** W202-8.

BAIN, J., PLATER, L., ELLIOTT, M., SHPIRO, N., HASTIE, C. J., MCLAUCHLAN, H., KLEVERNIC, I., ARTHUR, J. S., ALESSI, D. R. & COHEN, P. 2007. The selectivity of protein kinase inhibitors: a further update. *Biochem J,* 408**,** 297-315.

BASHIR, S. & PAESHUYSE, J. 2020. Construction of Antibody Phage Libraries and Their Application in Veterinary Immunovirology. *Antibodies (Basel),* 9.

BASSE, M. J., BETZI, S., BOURGEAS, R., BOUZIDI, S., CHETRIT, B., HAMON, V., MORELLI, X. & ROCHE, P. 2013. 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions. *Nucleic Acids Res,* 41**,** D824-7.

BATES, A. & POWER, C. A. 2019. David vs. Goliath: The Structure, Function, and Clinical Prospects of Antibody Fragments. *Antibodies,* 8.

BAUD, V. & KARIN, M. 2009. Is NF-kappaB a good target for cancer therapy? Hopes and pitfalls. *Nat Rev Drug Discov,* 8**,** 33-40.

BERG, J. M., TYMOCZKO, J. L. & STRYER, L. 2002. *Most Molecular-Motor Proteins Are Members of the P-Loop NTPase Superfamily,* New York, W H Freeman.

BIRD, R. E. & WEBB WALKER, B. 1991. Single chain antibody variable regions. *Trends Biotechnology,* 9**,** 132-137.

BUTLER, S. A., KHANLIAN, S. A. & COLE, L. A. 2001. Detection of Early Pregnancy Forms of Human Chorionic Gonadotropin by Home Pregnancy Test Devices. *Clinical Chemistry,* 47**,** 2131-2136.

CLAUSSON, C. M., ALLALOU, A., WEIBRECHT, I., MAHMOUDI, S., FARNEBO, M., LANDEGREN, U., WAHLBY, C. & SODERBERG, O. 2011. Increasing the dynamic range of in situ PLA. *Nat Methods,* 8**,** 892-3.

COLE, J. L., LARY, J. W., P. MOODY, T. & LAUE, T. M. 2008. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. *Biophysical Tools for Biologists, Volume One: In Vitro Techniques.*

COX, A. D., FESIK, S. W., KIMMELMAN, A. C., LUO, J. & DER, C. J. 2014. Drugging the undruggable RAS: Mission possible? *Nat Rev Drug Discov,* 13**,** 828-51.

CRUZ-MIGONI, A., CANNING, P., QUEVEDO, C. E., BATAILLE, C. J. R., BERY, N., MILLER, A., RUSSELL, A. J., PHILLIPS, S. E. V., CARR, S. B. & RABBITTS, T. H. 2019. Structure-based development of new RAS-effector inhibitors from a combination of active and inactive RAS-binding compounds. *Proc Natl Acad Sci U S A,* 116**,** 2545-2550.

CWIRLA, S. E., PETERS, E. A., BARRETT, R. W. & DOWER, W. J. 1990. Peptides on phage: a vast library of peptides for identifying ligands. *Proc Natl Acad Sci U S A,* 87**,** 6378-82.
DANG, C. V. 2012. MYC on the path to cancer. *Cell,* 149**,** 22-35.
DANG, C. V., REDDY, E. P., SHOKAT, K. M. & SOUCEK, L. 2017. Drugging the 'undruggable' cancer targets. *Nature Reviews Cancer,* 17**,** 502-508.
DEVLIN, J. J., PANGANIBAN, L. C. & DEVLIN, P. E. 1990. Random peptide libraries: a source of specific protein binding molecules. *Science,* 249**,** 404-6.
DIAS-NETO, E., NUNES, D. N., GIORDANO, R. J., SUN, J., BOTZ, G. H., YANG, K., SETUBAL, J. C., PASQUALINI, R. & ARAP, W. 2009. Next-generation phage display: integrating and comparing available molecular tools to enable cost-effective high-throughput analysis. *PLoS One,* 4**,** e8338.
DONALDSON, J. G. 2015. Immunofluorescence Staining. *Current Protocols in Cell Biology,* 69**,** 4.3.1-4.3.7.
ELION, J., BERG, P. E., LAPOUMEROULIE, C., TRABUCHET, G., MITTELMAN, M., KRISHNAMOORTHY, R., SCHECHTER, A. N. & LABIE, D. 1992. DNA sequence variation in a negative control region 5' to the beta- globin gene correlates with the phenotypic expression of the beta s mutation. *Blood,* 79**,** 787-792.
ENGVALL, E. 1980. [28] Enzyme immunoassay ELISA and EMIT. *Methods in Enzymology.* Academic Press.
FERNANDES, J. C. 2018. Therapeutic application of antibody fragments in autoimmune diseases: current state and prospects. *Drug Discov Today,* 23**,** 1996-2002.
FOTIS, C., ANTORANZ, A., HATZIAVRAMIDIS, D., SAKELLAROPOULOS, T. & ALEXOPOULOS, L. G. 2018. Network-based technologies for early drug discovery. *Drug Discov Today,* 23**,** 626-635.
FREDRICKSSON, S., GULLBERG, M., JARVIUS, J., OLSSON, C., PIETRAS, K., GUSTAFSDOTTIR, S. M., OSTMAN, A. & LANDEGREN, U. 2002. Protein detection using proximity-dependent DNA ligation assays. *Nature Biotechnology,* 20**,** 473-477.
FRIDMAN, J. S. & LOWE, S. W. 2003. Control of apoptosis by p53. *Oncogene,* 22**,** 9030-9040.
FRITZSCHE, F. R., DAHL, E., DANKOF, A., BURKHARDT, M., PAHL, S., PETERSEN, I., DIETEL, M. & KRISTIANSEN, G. 2007. Expression of ARG2 in non small cell lung cancer. *Histol Histopathol,* 22**,** 703-708.
FRITZSCHE, F. R., DAHL, E., PAHL, S., BURKHARDT, M., LUO, J., MAYORDOMO, E., GANSUKH, T., DANKOF, A., KNUECHEL, R., DENKERT, C., WINZER, K. J., DIETEL, M. & KRISTIANSEN, G. 2006. Prognostic relevance of AGR2 expression in breast cancer. *Clin Cancer Res,* 12**,** 1728-34.
FUJIWARA, Y., FUJIWARA, K., GODA, N., IWAYA, N., TENNO, T., SHIRAKAWA, M. & HIROAKI, H. 2011. Structure and function of the N-terminal nucleolin binding domain of nuclear valosin-containing

protein-like 2 (NVL2) harboring a nucleolar localization signal. *J Biol Chem,* 286**,** 21732-41.
GAI, D., LI, D., FINKIELSTEIN, C. V., OTT, R. D., TANEJA, P., FANNING, E. & CHEN, X. S. 2004. Insights into the oligomeric states, conformational changes, and helicase activities of SV40 large tumor antigen. *J Biol Chem,* 279**,** 38952-9.
GLEAVE, E. S., SCHMIDT, H. & CARTER, A. P. 2014. A structural analysis of the AAA+ domains in Saccharomyces cerevisiae cytoplasmic dynein. *J Struct Biol,* 186**,** 367-75.
GORYNIA, S., BANDEIRAS, T. M., PINHO, F. G., MCVEY, C. E., VONRHEIN, C., ROUND, A., SVERGUN, D. I., DONNER, P., MATIAS, P. M. & CARRONDO, M. A. 2011. Structural and functional insights into a dodecameric molecular machine - the RuvBL1/RuvBL2 complex. *J Struct Biol,* 176**,** 279-91.
GRAUDENS, E., BOULANGER, V., MOLLARD, C., MARIAGE-SAMSON, R., BARLET, X., GREMY, G., COUILLAULT, C., LAJEMI, M., PIATIER-TONNEAU, D., ZABORSKI, P., EVENO, E., AUFFRAY, C. & IMBEAUD, S. 2006. Deciphering cellular states of innate tumor drug responses. *Genome Biol,* 7**,** R19.
GULLBERG, M., GUSTAFSDOTTIR, S. M., SCHALLMEINER, E., JARVIUS, J., BJARNEGARD, M., BETSHOLTZ, C., LANDEGREN, U. & FREDRIKSSON, S. 2004. Cytokine detection by antibody-based proximity ligation. *Proc Natl Acad Sci U S A,* 101**,** 8420-4.
HANSON, P. I. & WHITEHEART, S. W. 2005. AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol,* 6**,** 519-29.
HANTSCHEL, O., GREBIEN, F. & SUPERTI-FURGA, G. 2011. Targeting allosteric regulatory modules in oncoproteins: "drugging the undruggable". *Oncotarget,* 2**,** 828-9.
HEALY, A. R., HOUSTON, D. R., REMNANT, L., HUART, A.-S., BRYCHTOVA, V., MASLON, M. M., MEERS, O., MULLER, P., KREJCI, A., BLACKBURN, E. A., VOJTESEK, B., HERNYCHOVA, L., WALKINSHAW, M. D., WESTWOOD, N. J. & HUPP, T. R. 2015. Discovery of a novel ligand that modulates the protein–protein interactions of the AAA+ superfamily oncoprotein reptin. *Chemical Science,* 6**,** 3109-3116.
HRSTKA, R., NENUTIL, R., FOURTOUNA, A., MASLON, M. M., NAUGHTON, C., LANGDON, S., MURRAY, E., LARIONOV, A., PETRAKOVA, K., MULLER, P., DIXON, M. J., HUPP, T. R. & VOJTESEK, B. 2010. The pro-metastatic protein anterior gradient-2 predicts poor prognosis in tamoxifen-treated breast cancers. *Oncogene,* 29**,** 4838-47.
HUPP, T. R., MEEK, D. W., MIDGLEY, C. A. & LANE, D. P. 1992. Regulation of the specific DNA binding function of p53. *Cell,* 71**,** 875-886.
IIZUKA, N., TSUNEDOMI, R., TAMESA, T., OKADA, T., SAKAMOTO, K., HAMAGUCHI, T., YAMADA-OKABE, H., MIYAMOTO, T., UCHIMURA, S., HAMAMOTO, Y. & OKA, M. 2006. Involvement of c-myc-regulated genes in hepatocellular carcinoma related to genotype-C hepatitis B virus. *J Cancer Res Clin Oncol,* 132**,** 473-81.

JAIN, S., ARESU, L., COMAZZI, S., SHI, J., WORRALL, E., CLAYTON, J., HUMPHRIES, W., HEMMINGTON, S., DAVIS, P., MURRAY, E., LIMENEH, A. A., BALL, K., RUCKOVA, E., MULLER, P., VOJTESEK, B., FAHRAEUS, R., ARGYLE, D. & HUPP, T. R. 2016. The Development of a Recombinant scFv Monoclonal Antibody Targeting Canine CD20 for Use in Comparative Medicine. *PLoS One,* 11**,** e0148366.

JESPERSEN, N. & BARBAR, E. 2020. Emerging Features of Linear Motif-Binding Hub Proteins. *Trends Biochem Sci,* 45**,** 375-384.

JUBB, H., HIGUERUELO, A. P., WINTER, A. & BLUNDELL, T. L. 2012. Structural biology and drug discovery for protein-protein interactions. *Trends Pharmacol Sci,* 33**,** 241-8.

KATTA, V. & CHAIT, B. T. 1993. Hydrogen-Deuterium Exchange Electrospray-Ionization Mass-Spectrometry - a Method for Probing Protein Conformational-Changes in Solution. *Journal of the American Chemical Society,* 115**,** 6317-6321.

KATZEN, F. 2007. Gateway® recombinational cloning: a biological operating system. *Expert Opinion on Drug Discovery,* 2**,** 571-589.

KESKIN, O. 2007. Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: a case study of antibodies. *BMC Struct Biol,* 7**,** 31.

KESSLER, D., GMACHL, M., MANTOULIDIS, A., MARTIN, L. J., ZOEPHEL, A., MAYER, M., GOLLNER, A., COVINI, D., FISCHER, S., GERSTBERGER, T., GMASCHITZ, T., GOODWIN, C., GREB, P., HARING, D., HELA, W., HOFFMANN, J., KAROLYI-OEZGUER, J., KNESL, P., KORNIGG, S., KOEGL, M., KOUSEK, R., LAMARRE, L., MOSER, F., MUNICO-MARTINEZ, S., PEINSIPP, C., PHAN, J., RINNENTHAL, J., SAI, J., SALAMON, C., SCHERBANTIN, Y., SCHIPANY, K., SCHNITZER, R., SCHRENK, A., SHARPS, B., SISZLER, G., SUN, Q., WATERSON, A., WOLKERSTORFER, B., ZEEB, M., PEARSON, M., FESIK, S. W. & MCCONNELL, D. B. 2019. Drugging an undruggable pocket on KRAS. *Proc Natl Acad Sci U S A,* 116**,** 15823-15829.

KHOLODENKO, R., KALINOVSKY, D., DORONIN, I., PONOMAREV, E. & KHOLODENKO, I. 2017. Antibody Fragments as Potential Biopharmaceuticals for Cancer Therapy: Success and Limitations. *Current medicinal chemistry,* 26.

KIM, D., YOU, E. & RHEE, S. 2012. Dynein regulates cell migration depending on substrate rigidity. *Int J Mol Med,* 29**,** 440-6.

KIM, J. H., CHOI, H. J., KIM, B., KIM, M. H., LEE, J. M., KIM, I. S., LEE, M. H., CHOI, S. J., KIM, K. I., KIM, S. I., CHUNG, C. H. & BAEK, S. H. 2006. Roles of sumoylation of a reptin chromatin-remodelling complex in cancer metastasis. *Nat Cell Biol,* 8**,** 631-9.

KIM, J. H., KIM, B., CAI, L., CHOI, H. J., OHGI, K. A., TRAN, C., CHEN, C., CHUNG, C. H., HUBER, O., ROSE, D. W., SAWYERS, C. L., ROSENFELD, M. G. & BAEK, S. H. 2005. Transcriptional regulation of a metastasis suppressor gene by Tip60 and beta-catenin complexes. *Nature,* 434**,** 921-926.

KNIGHT, Z. A. & SHOKAT, K. M. 2005. Features of selective kinase inhibitors. *Chem Biol,* 12**,** 621-37.
KOHLWAY, A., LUO, D., RAWLING, D. C., DING, S. C. & PYLE, A. M. 2013. Defining the functional determinants for RNA surveillance by RIG-I. *EMBO Rep,* 14**,** 772-9.
KUMAR, M., GOUW, M., MICHAEL, S., SAMANO-SANCHEZ, H., PANCSA, R., GLAVINA, J., DIAKOGIANNI, A., VALVERDE, J. A., BUKIROVA, D., CALYSEVA, J., PALOPOLI, N., DAVEY, N. E., CHEMES, L. B. & GIBSON, T. J. 2020. ELM-the eukaryotic linear motif resource in 2020. *Nucleic Acids Res,* 48**,** D296-D306.
LEIPE, D. D., WOLF, Y. I., KOONIN, E. V. & ARAVIND, L. 2002. Classification and evolution of P-loop GTPases and related ATPases. *J Mol Biol,* 317**,** 41-72.
LIU, D., RUDLAND, P. S., SIBSON, D. R., PLATT-HIGGINS, A. & BARRACLOUGH, R. 2005. Human homologue of cement gland protein, a novel metastasis inducer associated with breast carcinomas. *Cancer Res,* 65**,** 3796-805.
LIU, J. K. 2014. The history of monoclonal antibody development - Progress, remaining challenges and future innovations. *Ann Med Surg (Lond),* 3**,** 113-6.
LIU, Y., WAN, W. Z., LI, Y., ZHOU, G. L. & LIU, X. G. 2017. Recent development of ATP-competitive small molecule phosphatidylinostitol-3-kinase inhibitors as anticancer agents. *Oncotarget,* 8**,** 7181-7200.
LOPEZ-PERROTE, A., MUNOZ-HERNANDEZ, H., GIL, D. & LLORCA, O. 2012. Conformational transitions regulate the exposure of a DNA-binding domain in the RuvBL1-RuvBL2 complex. *Nucleic Acids Res,* 40**,** 11086-99.
MAO, Y. Q. & HOURY, W. A. 2017. The Role of Pontin and Reptin in Cellular Physiology and Cancer Etiology. *Front Mol Biosci,* 4**,** 58.
MASLON, M. M., HRSTKA, R., VOJTESEK, B. & HUPP, T. R. 2010. A Divergent Substrate-Binding Loop within the Pro-oncogenic Protein Anterior Gradient-2 Forms a Docking Site for Reptin. *Journal of Molecular Biology,* 404**,** 418-438.
MATIAS, P. M., GORYNIA, S., DONNER, P. & CARRONDO, M. A. 2006. Crystal structure of the human AAA+ protein RuvBL1. *J Biol Chem,* 281**,** 38918-29.
MATOCHKO, W. L., CHU, K., JIN, B., LEE, S. W., WHITESIDES, G. M. & DERDA, R. 2012. Deep sequencing analysis of phage libraries using Illumina platform. *Methods,* 58**,** 47-55.
MCCULLOUGH, B. J., KALAPOTHAKIS, J., EASTWOOD, H., KEMPER, P., MACMILLAN, D., TAYLOR, K., DORIN, J. & BARRAN, P. E. 2008. Development of an Ion Mobility Quadrupole Time of Flight Mass Spectrometer. *Analytical Chemistry,* 80**,** 6336-6344.
MILLER, M. L., JENSEN, L. J., DIELLA, F., JORGENSEN, C., TINTI, M., LI, L., HSIUNG, M., PARKER, S. A., BORDEAUX, J., SICHERITZ-PONTEN, T., OLHOVSKY, M., PASCULESCU, A., ALEXANDER, J., KNAPP, S., BLOM, N., BORK, P., LI, S., CESARENI, G., PAWSON, T., TURK, B. E., YAFFE, M. B., BRUNAK, S. & LINDING, R. 2008.

Linear motif atlas for phosphorylation-dependent signaling. *Sci Signal,* 1**,** ra2.
MITCHELL, W. & MATSUMOTO, S. 2011. Large-scale integrated super-computing platform for next generation virtual drug discovery. *Curr Opin Chem Biol,* 15**,** 553-9.
MOZAFFARI-JOVIN, S., WANDERSLEBEN, T., SANTOS, K. F., WILL, C. L., LUHRMANN, R. & WAHL, M. C. 2013. Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. *Science,* 341**,** 80-4.
NANO, N. & HOURY, W. A. 2013. Chaperone-like activity of the AAA+ proteins Rvb1 and Rvb2 in the assembly of various complexes. *Philos Trans R Soc Lond B Biol Sci,* 368**,** 20110399.
NANO, N., UGWU, F., SERAPHIM, T. V., LI, T., AZER, G., ISAAC, M., PRAKESCH, M., BARBOSA, L. R. S., RAMOS, C. H. I., DATTI, A. & HOURY, W. A. 2020. Sorafenib as an Inhibitor of RUVBL2. *Biomolecules,* 10.
ODELL, I. D. & COOK, D. 2013. Immunofluorescence techniques. *J Invest Dermatol,* 133**,** e4.
ONG, S. E., BLAGOEV, B., KRATCHMAROVA, I., KRISTENSEN, D. B., STEEN, H., PANDEY, A. & MANN, M. 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics,* 1**,** 376-386.
PERRI, F., PISCONTI, S. & DELLA VITTORIA SCARPATI, G. 2016. P53 mutations and cancer: a tight linkage. *Annals of Translational Medicine,* 4**,** 522-522.
PETUKHOV, M., DAGKESSAMANSKAJA, A., BOMMER, M., BARRETT, T., TSANEVA, I., YAKIMOV, A., QUÉVAL, R., SHVETSOV, A., KHODORKOVSKIY, M., KÄS, E. & GRIGORIEV, M. 2012. Large-Scale Conformational Flexibility Determines the Properties of AAA+ TIP49 ATPases. *Structure,* 20**,** 1321-1331.
PIETENPOL, J. A., T., T., THIAGALINGAM, S., EL-DEIRY, W. S., KINZLER, K. W. & VOGELSTEIN, B. 1994. Sequence-specific transcriptional activation is essential for growth suppression by p53. *Proc Natl Acad Sci U S A,* 91**,** 1998-2002.
PISA, R., CUPIDO, T., STEINMAN, J. B., JONES, N. H. & KAPOOR, T. M. 2019. Analyzing Resistance to Design Selective Chemical Inhibitors for AAA Proteins. *Cell Chem Biol,* 26**,** 1263-1273 e5.
POHLER, E., CRAIG, A. L., COTTON, J., LAWRIE, L., DILLON, J. F., ROSS, P., KERNOHAN, N. & HUPP, T. R. 2004. The Barrett's antigen anterior gradient-2 silences the p53 transcriptional response to DNA damage. *Mol Cell Proteomics,* 3**,** 534-47.
QUEVAL, R., PAPIN, C., DALVAI, M., BYSTRICKY, K. & HUMBERT, O. 2014. Reptin and Pontin oligomerization and activity are modulated through histone H3 N-terminal tail interaction. *J Biol Chem,* 289**,** 33999-4012.
QUEVEDO, C. E., CRUZ-MIGONI, A., BERY, N., MILLER, A., TANAKA, T., PETCH, D., BATAILLE, C. J. R., LEE, L. Y. W., FALLON, P. S.,

TULMIN, H., EHEBAUER, M. T., FERNANDEZ-FUENTES, N., RUSSELL, A. J., CARR, S. B., PHILLIPS, S. E. V. & RABBITTS, T. H. 2018. Small molecule inhibitors of RAS-effector protein interactions derived using an intracellular antibody fragment. *Nat Commun,* 9**,** 3169.

RAMASAMY, S., BENNET, D. & KIM, S. 2014. Drug and bioactive molecule screening based on a bioelectrical impedance cell culture platform. *Int J Nanomedicine,* 9**,** 5789-809.

RAY-COQUARD, I., BLAY, J.-Y., ITALIANO, A., LE CESNE, A., PENEL, N., ZHI, J., HEIL, F., RUEGER, R., GRAVES, B., DING, M., GEHO, D., MIDDLETON, S. A., VASSILEV, L. T., NICHOLS, G. L. & BUI, B. N. 2012. Effect of the MDM2 antagonist RG7112 on the P53 pathway in patients with MDM2-amplified, well-differentiated or dedifferentiated liposarcoma: an exploratory proof-of-mechanism study. *The Lancet Oncology,* 13**,** 1133-1140.

RAYCROFT, L., WU, H. Y. & LOZANO, G. 1990. Transcriptional activation by wild-type but not transforming mutants of the p53 anti-oncogene. *Science,* 248.

REINER, M.-O., PILARSKY, C., GERHARDT, J., GRUTZMANN, R., FRITZSCHE, F. R., BAHRA, M., WEICHERT, W. & KRISTIANSEN, G. 2009. Prognostic significance of AGR2 in pancreatic ductal adenocarcinoma. *Histol Histopathol,* 24**,** 1121-1128.

REMNANT, L., COUFALOVA, D., HERNYCHOVA, L., MULLER, P., HEALY, A., KANNAN, S., WESTWOOD, N., VERMA, C. S., VOJTESEK, B., HUPP, T. R. & HOUSTON, D. R. 2019. An inter-subunit protein-peptide interface that stabilizes the specific activity and oligomerization of the AAA+ chaperone Reptin. *J Proteomics,* 199**,** 89-101.

RENA, G., BAIN, J., ELLIOTT, M. & COHEN, P. 2004. D4476, a cell-permeant inhibitor of CK1, suppresses the site-specific phosphorylation and nuclear exclusion of FOXO1a. *EMBO Rep,* 5**,** 60-5.

ROSENBAUM, J., BAEK, S. H., DUTTA, A., HOURY, W. A., HUBER, O., HUPP, T. R. & MATIAS, P. M. 2013. The emergence of the conserved AAA+ ATPases Pontin and Reptin on the signaling landscape. *Sci Signal,* 6**,** mr1.

ROUSSEAU, B., MENARD, L., HAURIE, V., TARAS, D., BLANC, J. F., MOREAU-GAUDRY, F., METZLER, P., HUGUES, M., BOYAULT, S., LEMIERE, S., CANRON, X., COSTET, P., COLE, M., BALABAUD, C., BIOULAC-SAGE, P., ZUCMAN-ROSSI, J. & ROSENBAUM, J. 2007. Overexpression and role of the ATPase and putative DNA helicase RuvB-like 2 in human hepatocellular carcinoma. *Hepatology,* 46**,** 1108-18.

SAIDA, F., UZAN, M., ODAERT, B. & BONTEMS, F. 2006. Expression of highly toxic genes in E-coli: Special strategies and genetic tools. *Current Protein & Peptide Science,* 7**,** 47-56.

SANCHEZ-CARBAYO, M., SOCCI, N. D., LOZANO, J., SAINT, F. & CORDON-CARDO, C. 2006. Defining molecular profiles of poor

outcome in patients with invasive bladder cancer using oligonucleotide microarrays. *J Clin Oncol,* 24**,** 778-89.
SCOTT, A., CHUNG, H. Y., GONCIARZ-SWIATEK, M., HILL, G. C., WHITBY, F. G., GASPAR, J., HOLTON, J. M., VISWANATHAN, R., GHAFFARIAN, S., HILL, C. P. & SUNDQUIST, W. I. 2005. Structural and mechanistic studies of VPS4 proteins. *EMBO J,* 24**,** 3658-69.
SCOTT, A. M., WOLCHOK, J. D. & OLD, L. J. 2012. Antibody therapy of cancer. *Nat Rev Cancer,* 12**,** 278-87.
SCOTT, J. K. & SMITH, G. P. 1990. Searching for peptide ligands with an epitope library. *Science,* 249**,** 386-90.
SMITH, G. P. 1985. Filamentous Fusion Phage - Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science,* 228**,** 1315-1317.
SNIDER, J., THIBAULT, G. & HOURY, W. A. 2008. The AAA+ superfamily of functionally diverse proteins. *Genome Biol,* 9**,** 216.
STINSON, B. M., NAGER, A. R., GLYNN, S. E., SCHMITZ, K. R., BAKER, T. A. & SAUER, R. T. 2013. Nucleotide binding and conformational switching in the hexameric ring of a AAA+ machine. *Cell,* 153**,** 628-39.
TALANTOV, D., MAZUMDER, A., YU, J. X., BRIGGS, T., JIANG, Y., BACKUS, J., ATKINS, D. & WANG, Y. 2005. Novel genes associated with malignant melanoma but not benign melanocytic lesions. *Clin Cancer Res,* 11**,** 7234-42.
TARAZONA, M. P. & SAIZ, E. 2003. Combination of SEC/MALS experimental procedures and theoretical analysis for studying the solution properties of macromolecules. *Journal of Biochemical and Biophysical Methods,* 56**,** 95-116.
TAYLOR, P., BLACKBURN, E., SHENG, Y. G., HARDING, S., HSIN, K. Y., KAN, D., SHAVE, S. & WALKINSHAW, M. D. 2008. Ligand discovery and virtual screening using the program LIDAEUS. *Br J Pharmacol,* 153 Suppl 1**,** S55-67.
THOMPSON, A., SCHAFER, J., KUHN, K., KIENLE, S., SCHWARZ, J., SCHMIDT, G., NEUMANN, T., JOHNSTONE, R., MOHAMMED, A. K. & HAMON, C. 2003. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem,* 75**,** 1895-904.
TOMPA, P., DAVEY, N. E., GIBSON, T. J. & BABU, M. M. 2014. A million peptide motifs for the molecular biologist. *Mol Cell,* 55**,** 161-9.
VALE, R. D. 2000. AAA proteins: Lords of the ring. *Journal of Cell Biology,* 150**,** F13-F19.
VAN ROEY, K., UYAR, B., WEATHERITT, R. J., DINKEL, H., SEILER, M., BUDD, A., GIBSON, T. J. & DAVEY, N. E. 2014. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev,* 114**,** 6733-78.
VENTEICHER, A. S., MENG, Z., MASON, P. J., VEENSTRA, T. D. & ARTANDI, S. E. 2008. Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly. *Cell,* 132**,** 945-975.

WALLRABE, H. & PERIASAMY, A. 2005. Imaging protein molecules using FRET and FLIM microscopy. *Current Opinion in Biotechnology,* 16**,** 19-27.

WASSERMAN. 2020. *Cell Cycle and Cancer* [Online]. CuboCube: CuboCube. Available: http://www.cubocube.com/dashboard.php?a=1642&b=1691&c=1 [Accessed 2020].

WEIKL, T. R. & PAUL, F. 2014. Conformational selection in protein binding and function. *Protein Science,* 23**,** 1508-1518.

WU, K. J., LEI, P. M., LIU, H., WU, C., LEUNG, C. H. & MA, D. L. 2019. Mimicking Strategy for Protein-Protein Interaction Inhibitor Discovery by Virtual Screening. *Molecules,* 24.

ZHANG, J. S., GONG, A., CHEVILLE, J. C., SMITH, D. I. & YOUNG, C. Y. 2005. AGR2, an androgen-inducible secretory protein overexpressed in prostate cancer. *Genes Chromosomes Cancer,* 43**,** 249-59.

Chapter 9 Appendices

Development of a fluorescent monoclonal antibody-based assay to measure the allosteric effects of synthetic peptides on self-oligomerization of AGR2 protein.
Gray TA, Murray E, Nowicki MW, **Remnant L**, Scherl A, Muller P, Vojtesek B, Hupp TR.
Protein Sci. 2013 Sep;22(9):1266-78. doi: 10.1002/pro.2299. Epub 2013 Jul 25.
PMID:23780840

Discovery of a novel ligand that modulates the protein-protein interactions of the AAA+ superfamily oncoprotein reptin.
Healy AR, Houston DR, **Remnant L**, Huart AS, Brychtova V, Maslon MM, Meers O, Muller P, Krejci A, Blackburn EA, Vojtesek B, Hernychova L, Walkinshaw MD, Westwood NJ, Hupp TR.
Chem Sci. 2015 May 1;6(5):3109-3116. doi: 10.1039/c4sc03885a. Epub 2015 Mar 20.
PMID:28706685

An inter-subunit protein-peptide interface that stabilizes the specific activity and oligomerization of the AAA+ chaperone Reptin.
Remnant L, Coufalova D, Hernychova L, Muller P, Healy A, Kannan S, Westwood N, Verma CS, Vojtesek B, Hupp TR, Houston DR.
J Proteomics. 2019 May 15;199:89-101. doi: 10.1016/j.jprot.2019.02.012. Epub 2019 Mar 9.
PMID:30862565

9.1 Development of a fluorescent monoclonal antibody-based assay to measure the allosteric effects of synthetic peptides on self-oligomerization of AGR2 protein

- Fig. 8 The effects of AGR2 oligomerization on its protein interaction activity
    - Expression of Reptin
    - AGR2-Reptin protein interaction assay

9.2 Discovery of a novel ligand that modulates the protein-protein interactions of the AAA+ superfamily oncoprotein Reptin.

- Fig. 1 Identification of small molecules targeting reptin using an in-silico screening program
    - HDX data
- Fig. 2 The effect of Liddean on the oligomerization dynamics of reptin
    - SDS-PAGE higher order structures
    - Reptin-HSP90 interaction assay
- Fig. 3 Discovery of new Liddean-dependent interaction motifs for reptin
    - Peptide phage display
    - Sequencing analysis
    - MEME
- Fig 4. Identification and validation of p53 as a novel reptin interacting protein
    - Motif search
    - P53 discovery
    - Binding of Reptin and ARG2 to p53
    - Changes in Reptin-POI complex with addition of Liddean
- Fig 5. Fine mapping of the dominant linear peptide docking site of reptin on p53
    - Peptide mapping of Reptin binding site on p53
- Fig 6. The effects of Liddean on reptin–pontin interactions in cell models using the proximity ligation assay
    - PLA Reptin-Pontin
    - Chemical fractionation testing levels of Reptin and Pontin on Liddean addition

9.3 An inter-subunit protein-peptide interface that stabilizes the specific activity and oligomerization of the AAA+ chaperone Reptin.

- Fig 1. Structure of Reptin
    - Self-binding peptide ELISA
    - Y340 localization
- Fig 2. Impact of ligands on the protein-binding function of Reptin
    - Peptide and protein binding assays
- Fig 3. The impact of mutation on Reptin quaternary structure and specific activity
    - Gel filtration of WT and mutants
    - Reptin WT and mutants binding to AGR2
- Fig 4. A model describing the specific activity of Reptin as a function of its oligomeric state
- Fig 5. Hydrogen-Deuterium exchange mass spectrometry of Reptin in differing conformational states
    - Creation of Y340A mutant protein
    - Creation of oligomeric proteins for analysis with HDX

THE PROTEIN SOCIETY

METHODS AND APPLICATIONS

Development of a fluorescent monoclonal antibody-based assay to measure the allosteric effects of synthetic peptides on self-oligomerization of AGR2 protein

Terry A. Gray,[1] Euan Murray,[1,2] Matthew W. Nowicki,[1] Lucy Remnant,[1] Alexander Scherl,[3] Petr Muller,[2] Borek Vojtesek,[2] and Ted R. Hupp[1]*

[1]Institute of Genetics and Molecular Medicine, Cell Signaling Unit, University of Edinburgh, United Kingdom
[2]Regional Centre for Applied Molecular Oncology, Masaryk Memorial Cancer Institute, 656 53 Brno, Czech Republic
[3]Dept. of Human Protein Sciences, Faculty of Medicine, University of Geneva, Geneva, Switzerland

Received 2 April 2013; Accepted 10 June 2013
DOI: 10.1002/pro.2299
Published online 18 June 2013 proteinscience.org

Abstract: Many regulatory proteins are homo-oligomeric and designing assays that measure self-assembly will provide novel approaches to study protein allostery and screen for novel small molecule modulators of protein interactions. We present an assay to begin to define the biochemical determinants that regulate dimerization of the cancer-associated oncoprotein AGR2. A two site-sandwich microtiter assay (^{2S}MTA) was designed using a DyLight800-labeled monoclonal antibody that binds to an epitope in AGR2 to screen for synthetic self-peptides that might regulate dimer stability. Peptides derived from the intrinsically disordered N-terminal region of AGR2 increase *in trans* oligomer stability as defined using the ^{2S}MTA assay. A DSS-crosslinking assay that traps the AGR2 dimer through K95-K95 adducts confirmed that Δ45-AGR2 was a more stable dimer using denaturing gel electrophoresis. A titration of wt-AGR2, Δ45-AGR2 (more stable dimer), and monomeric AGR2^{E60A} revealed that Δ45-AGR2 was more active in binding to Reptin than either wt-AGR2 or the AGR2^{E60A} mutant. Our data have defined a functional role for the AGR2 dimer in the binding to its most well characterized interacting protein, Reptin. The ability to regulate AGR2 oligomerization *in trans* opens the possibility for developing small molecules that regulate its' biochemical activity as potential cancer therapeutics. The data also highlight the utility of this oligomerization assay to screen chemical libraries for ligands that could regulate AGR2 dimer stability and its' oncogenic potential.

Keywords: oligomerization; allostery; protein interactions; monoclonal antibody

Introduction

The discovery of protein–protein interactions is a key, fundamental goal in life sciences research and protein–protein interactions are emerging as compelling landscapes in the drug discovery field.[1,2] The fundamental nature of a protein–protein interaction

Grant sponsors: Czech Science Foundation P301/11/1678; European Regional Development Fund and the State Budget of the Czech Republic (RECAMO, CZ.1.05/2.1.00/03.0101); BBSRC; CRUK.

*Correspondence to: Ted R. Hupp, Institute of Genetics and Molecular Medicine, Cell Signaling Unit, University of Edinburgh, UK. E-mail: ted.hupp@ed.ac.uk

Published by Wiley-Blackwell. © 2013 The Protein Society

has changed since the realization that intrinsically unstructured proteins form important components of a large number of protein–protein interactions.[3] This view of a protein-protein interaction provides approaches for using structural and computational sciences to develop lead molecules that disrupt a given protein–protein interaction.[4] In addition, the realization that proteins exist in large oligomeric signaling complexes provides a relatively large surface area for identifying protein-interaction hotspots that might be amenable to fragment-based lead molecule discovery.[5,6]

Structure-based lead molecule discovery remains a fundamental mechanism to exploit protein structure, develop small molecule drug leads, and manipulate allostery in proteins.[7] Amongst the many emerging models in the cancer field, for example, small molecules exist that stabilize the tetramer–dimer/monomer conversion of pyruvate kinase that has potential to regulate metabolism in cancer cells.[8,9] In addition, the structure of porphobilinogen synthase has led to models on the interconversion of octomers to hexamers through dimer assembly.[10] Small molecules have been developed that stabilize the hexameric form of the enzyme and inhibit enzyme activity[11] that highlight the ability to exploit the allosteric regulatory nature of many proteins by developing small molecule tools or drug leads.

One limitation of structure-based drug discovery in human disease-associated proteins is that a large proportion of polypeptide sequence information in higher eukaryotes is composed of intrinsically disordered regions that make some domains refractory to structural solutions.[12,13] These disordered regions can acquire structure by an induced-fit mechanism upon binding to their targets thus providing a specificity in the protein–protein interaction.[14] Proteins with disordered domains and oligomeric properties can form prototypes for developing innovative approaches to measure allostery in protein control.[15] The oncoprotein MDM2 has formed such a prototype model whose full-length properties have been refractory to structural biology solutions.[16–18] We had used MDM2 as a model to develop a nanoparticle binding assay using surface-enhanced RAMAN spectroscopy (SERS) that can measure the oligomerization state of small amounts of protein when bound to ligand thus providing a nano-sensing assay to measure protein allostery.[19] Although one methodological advantage of SERS as a nanoparticle binding assay is its use of small amounts of protein, a major limitation of SERS in measuring protein oligomerization is that it requires the target protein to be complexed to its ligand.

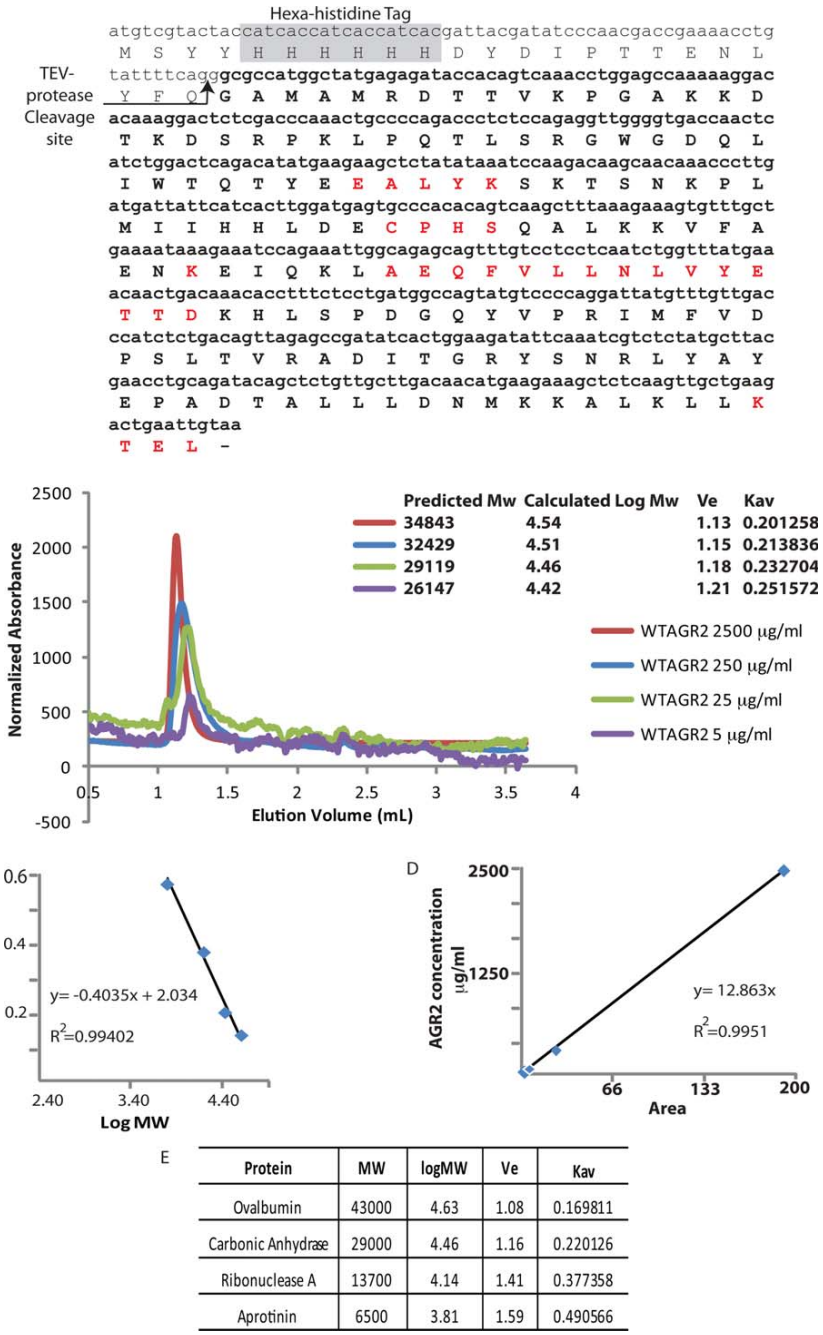
In this study, we present an alternative to the ligand-dependent SERS oligomerization assay by using a fluorescently labeled monoclonal antibody based microtiter assay that measures oligomerization of a ligand-free protein. The prototype protein used was the pro-oncogenic protein Anterior Gradient-2, that is an emerging cancer drug target in the endoplasmic-reticulum secretory system.[20–22] Anterior-Gradient-2 has an intrinsically disordered N-terminal region and the protein has recently been reported to be dimeric.[23] This fluorescent-based microtiter assay measured changes in AGR2 dimer stability in response to synthetic peptides derived from its intrinsically disordered domain *in trans*. This methodology provides a robust assay that can be used to screen chemical libraries for molecules that can alter AGR2 dimer stability and that can in turn be tested for inhibition of the oncogenic function of AGR2 in cells. In addition, this methodological concept is amenable in principle to other oligomeric proteins and can be used to screen for and isolate chemicals that alter oligomeric protein stability.

Results

Size exclusion chromatography demonstrates that AGR2 can exist as a homodimer

Gel-filtration size exclusion chromatography of recombinant mature $AGR2^{21–175}$ was used to define its oligomeric structure [Fig. 1(A)]. At highest concentrations (2.5 mg mL^{-1}; 136 μM), the protein eluted with a calculated molecular mass of 34.843 kDa [Fig. 1(B)], significantly greater than the theoretical monomeric molecular mass of 18252.1 kDa (Expasy Compute pI/MW tool, http://web.expasy.org/compute_pi/), more closely resembling a dimer structure relative to the theoretical mass of 36,504 kDa. The elution profiles of globular proteins used for calibration are summarized in Figure 1(C,E). By comparison, a prior analysis of $AGR2^{21–175}$ protein using SEC-MALLS gave a dimeric mass of 30.5 kDa from a protein with a monomeric mass of 17.8 kDa and predicted dimeric mass of 32.2 kDa.[23]

We diluted AGR2 protein from 136 μM down to 13.6, 1.36, and 0.27 μM prior to injection on the Sephadex-75 column to determine whether there is a concentration-dependence to dimerization [Fig. 1(B)]. AGR2 protein injected at a concentration of 13.6 μM eluted with an estimated mass of 32.429 kDa, protein injected at 1.36 μM exhibited a slower eluting species with an estimated mass of 29.119 kDa, and injection at a concentration of 0.27 μM displayed a mass of 26.147 kDa, suggesting that the protein can exist in a dimer-monomer equilibrium as it approaches predicted monomeric mass of 18.2 kDa at lower concentrations. The observed absorbance upon elution after each injection (214 nm; data not shown) corresponds to the starting concentration, as after integration of the peaks of each trace and plotting against the concentration, the values appear linear with



E

Protein	MW	logMW	Ve	Kav
Ovalbumin	43000	4.63	1.08	0.169811
Carbonic Anhydrase	29000	4.46	1.16	0.220126
Ribonuclease A	13700	4.14	1.41	0.377358
Aprotinin	6500	3.81	1.59	0.490566

Figure 1. Molecular mass of native AGR2$^{21–175}$ protein. (A) The amino acid sequence of the recombinant "mature" AGR2$^{21–175}$ protein lacking the N-terminal leader sequence, highlighting the: (i) hexa-histidine motif and TEV protease cleavage site in the tag, (ii) 60-EALYK-64 dimerization motif, (iii) 81-CPHS-84 thioredoxin motif, (iv) K95, (v) 101-AEQFVLLNLVYETTD-115 hydrophobic Reptin binding site, and (vi) degenerate 172-KTEL-175 ER retention site. (B) Gel filtration size exclusion chromatography of AGR2$^{21–175}$ protein, injected at the indicated concentrations, using an analytical Superdex PC 75 column. The data are plotted as relative absorbance to allow superimposition of the elution profiles. (C and E). Calibration of the Superdex PC 75 column was performed with the indicated gel filtration markers (GE) in SEC buffer (50 m*M* Tris pH 7.5, 200 m*M* NaCl. (D). Displays a plot of the peak area of the AGR2 peak (from B) as a function of AGR2 protein concentration at the time of injection, to highlight the linearity between protein absorbance upon elution (at 214 nm) and protein (concentration) injected. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$R^2 = 0.9951$ [Fig. 1(D)]. The traces in Figure 1(B) have been normalized for ease of visual comparison. These data are consistent with data suggesting that a stable dimeric species of AGR2 requires an *in vitro* concentration of the pure protein in the low μ*M* range.[23]

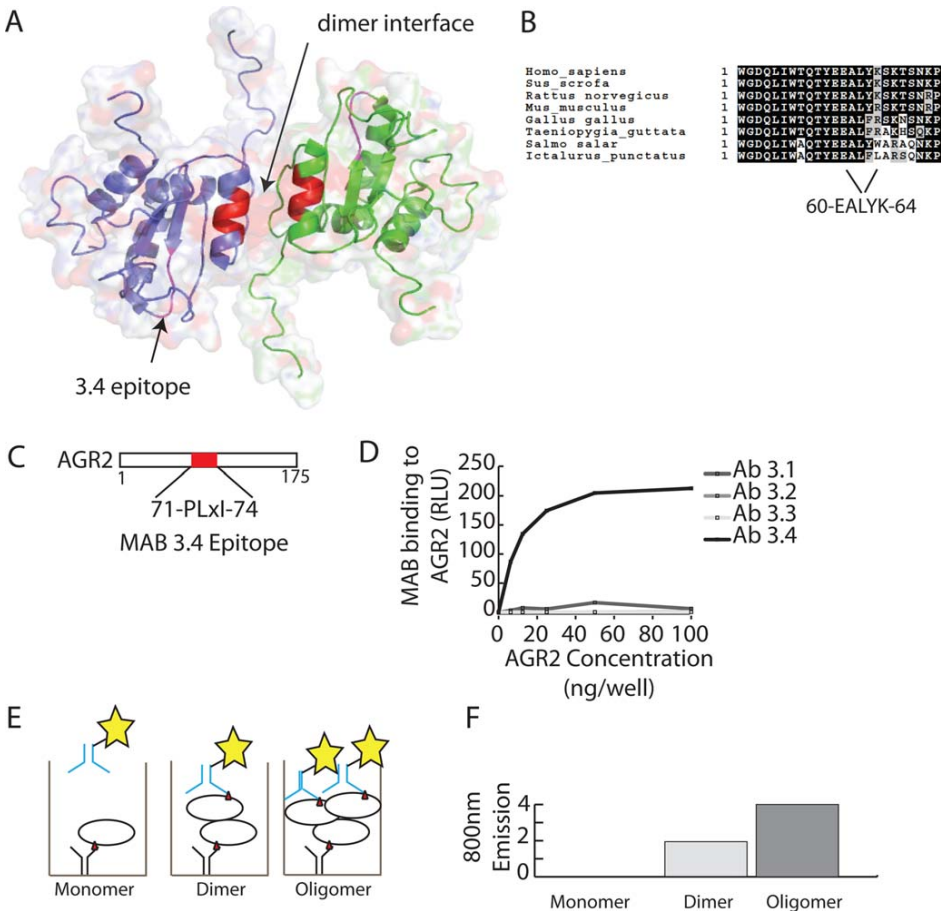


Figure 2. Localization of the epitope for MAB3.4 on AGR2 protein. (A) Cartoon of the dimeric structure of AGR2 (PDB; 2LNS;[23]) highlighting the dimer interface (B) and the MAB3.4 epitope (C). (D) ELISA-based assay analyzing the specific reactivity a panel of monoclonal antibodies raised against AGR3[25] toward AGR2. One of the AGR3-targetting monoclonal antibodies binds to AGR2 (MAB3.4). (E and F) Theory of analyzing and quantifying the oligomeric nature of AGR2 using the monoclonal antibody 3.4; (E) If AGR2 was monomeric and captured in the solid phase with MAB3.4, then the detection of this immune complex with the same MAB would not give a signal as there is not a second epitope (red triangle) present in the monomeric protein. However, a dimeric or oligomeric protein could be captured and detected with a monoclonal antibody that binds to the same epitope as there would be more than one epitope exposed in the immune complex. Although we can presume that AGR2 is a dimer in solution (as from Fig. 1 and from Patel *et al*.,[23]), this assay cannot distinguish between a dimer and higher oligomer, so we use oligomer throughout the text. (F) Predicted generic emission of an assay to measure AGR2 oligomerization using the fluorescently tagged MAB3.4. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Developing a quantitative microtiter assay to measure AGR2 oligomerization

It is not known whether the oligomeric (e.g., dimeric) structure of AGR2 is required for any of its protein-interaction functions.[23] To develop quantitative assays to measure AGR2 dimerization, we aimed to first determine whether a quantitative two-site "sandwich" microtiter assay ([2S]MTA) could be used to quantify oligomerization (e.g., dimerization). We had previously published a panel of monoclonal antibodies generated to the AGR2 orthologue, AGR3. Like AGR2,[24] AGR3 can mediate cisplatin resistance[25] in xenografts. Of these monoclonal antibodies,[25] one (MAB3.4), cross-reacted with AGR2 [Fig. 2(D)]. The AGR2 epitope recognized by MAB3.4 was fine mapped to a short linear 5 amino acid residue motif of 76-HHLED-80 [Fig. 2(C)];[25] that is out with the dimerization site [Fig. 2(A)] and therefore the antibody can be used in the [2S]MTA. The premise of the [2S]MTA is that the same immobilized MAB can both capture and detect the target protein only if the protein was oligomeric; for example, monomers cannot be detected by this assay [Fig. 2(E)]. Fluorescent labeling would allow quantitative detection of oligomers over monomers in real time [Fig. 2(F)]. As we cannot distinguish a dimer (based on gel filtration) from an oligomer using the [2S]MTA, we prefer to name the species we observe an oligomer.

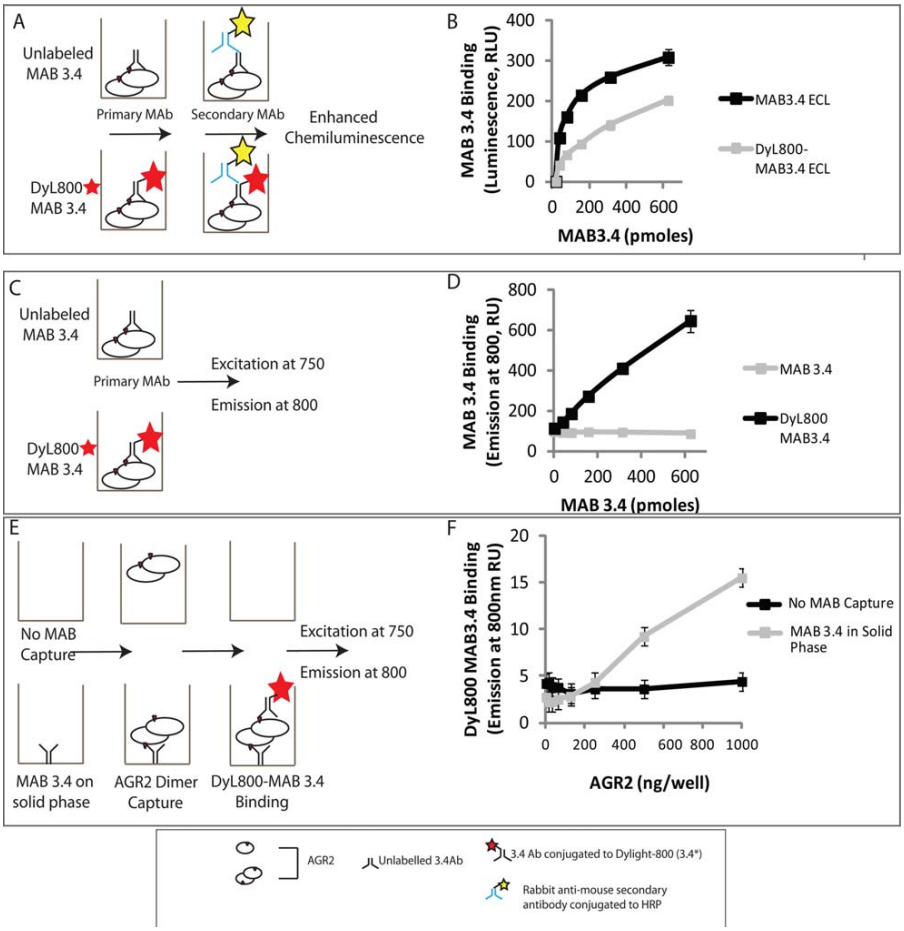


Figure 3. Developing a quantitative microtiter assay to measure AGR2 oligomerization. (A) Evaluation of the bioactivity of fluorescently labeled MAB3.4 in a luminescent-based ELISA. (B) MAB3.4 was left unconjugated or conjugated to DyL800 and after purification of the latter; the monoclonal antibodies were titrated into reactions containing AGR2 protein on the solid phase. The binding of MAB3.4 to AGR2 was measured using an anti-mouse IgG secondary antibody conjugated to peroxidase. The bioactivity of the monoclonal antibody (in relative luminescent units) is measured as a function of increasing MAB3.4 concentration. (C) Evaluation of the bioactivity of fluorescently labeled MAB3.4 in a fluorescence detection assay. (D) MAB3.4 was left unconjugated or conjugated to DyL800 and the monoclonal antibodies were titrated into reactions containing AGR2 protein on the solid phase. The bioactivity of the monoclonal antibody was measured as levels of emission at 800 nm a function of increasing MAB3.4 concentration. (E) Evaluation of the bioactivity of fluorescently labeled MAB3.4 to detect a potential AGR2 oligomer. (F) MAB3.4 was coated onto the solid phase and increasing amounts of oligomeric AGR2 were added to allow capture onto the solid phase. Fixed amounts of DyL800 MAB were added into reactions and the binding of DyL800-MAB3.4. The extent of oligomerization was quantified as levels of emission at 800 nm a function of increasing AGR2 protein concentration and is presented as an average from triplicate titrations. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Labeling the monoclonal antibody and optimization of the ^{2S}MTA

To determine whether MAB3.4 can be used to measure AGR2 oligomerization, it was used to capture AGR2 and then a DyLight800-(DyL800)-labeled version of MAB3.4 was used to detect the captured AGR2. If AGR2 was only a monomer, then there could not be any detection of AGR2 protein using DyL800-MAB3.4 in the solution phase. However, if AGR2 were oligomeric (e.g., dimeric or larger order assemblies), the epitope would, in principle, be presented to DyL800-MAB3.4 in the liquid phase and the extent of oligomerization would be proportional to the amount of DyL800-MAB3.4 bound [Fig. 2(E,F)].

To develop this ^{2S}MTA, we needed to first label the MAB3.4 with a fluorescent probe under conditions in which the labeled MAB retained functionality. Labeling of MAB3.4 was done with the far-infrared fluorophore DyL800. The labeling reaction proceeds through modification of primary amines and if an essential lysine in the complimentary determining region (CDR) was important for AGR2 binding, then

labeling might inactive the MAB [Fig. 3(A)]. A titration of MAB3.4 and DyL800-MAB3.4 reveals that the labeled antibody produces a binding activity in RLU ~1.8–2.5 lower fold lower than the unlabeled monoclonal antibody [Fig. 3(B)]. These data indicate that the DyL800-labeled monoclonal antibody retains significant affinity for its epitope in AGR2 protein and that binding could be quantified in a dose-dependent manner. The AGR2 protein-MAB interaction in Figure 3(B) measured the binding of the MAB3.4 or DyL800-MAB3.4 using peroxidase-labeled anti-mouse IgG secondary antibody and detection using chemiluminescence [Fig. 3(A)]. Following this, we evaluated binding using Licor far-infrared detection by excitation at 750 nm and fluorophore emission detected at 800 nm [Fig. 3(C)]. The DyL800-MAB3.4 binding to AGR2 adsorbed on the solid phase could be detected in a dose-dependent manner [Fig. 3(D)], relative to the unlabeled MAB3.4 that did not give rise to any signal. The binding reaction appeared more linear using DyL800-MAB3.4 in the direct excitation-emission assay [Fig. 3(D)] than in the indirect peroxidase-labeled secondary antibody [Fig. 3(B)]. There are two explanations for this phenomenon. First, the chemiluminescence assay [in Fig. 3(B)] uses "enhanced" enzymatic conversion by the antibody-peroxidase conjugate of substrate to create light that is quantified as a function of a fixed time. As a result, the data do not reflect a reaction rate but a final product accumulation. This "enhancement" of the primary signal may produce results that deviate from linearity. By contrast, the fluorescently conjugated monoclonal antibody emits a signal that is detected directly and the results are presumably more linear. Second, the indirect enhanced chemiluminescence assay incorporates an additional 75 min of incubation with secondary antibody, washings, and chemiluminescence substrate addition and this incorporates an unquantified effect of antibody off rate on the signal intensity. By contrast, after incubations with the fluorescently conjugated antibody, the reaction is washed rapidly, and fluorescence intensity is read immediately thus minimizing effects of the antibody off rate on the signal intensity. These data validate the DyL800-MAB3.4 excitation-emission assay as a bioactive probe. As such, we next evaluated whether DyL800-MAB3.4 could detect oligomeric forms of AGR2 in assay as a quantitative assay to measure oligomerization (oligomerization) of AGR2 protein [as outlined in Fig. 2(E)]. An outline of approach using DyL800-MAB3.4 to detect an oligomeric (potentially dimeric) species of AGR2 is outline in Figure 3(E). Reactions were assembled without or with adsorption of unlabeled MAB3.4 on the solid phase. Upon titration of DyL800-MAB3.4, a dose-dependent signal can be observed [Fig. 3(F)], suggesting that this assay can indeed measure oligomeric AGR2 protein.

Self-peptides that stabilize or disrupt the oligomer equilibrium

We next used the ^{2S}MTA to determine whether AGR2 oligomer stability can be attenuated using dimer interface peptides or indeed stabilized *in trans* by regulatory motifs from the intrinsically disordered region of the protein [Fig. 4(B)]. The predicted disorder of polypeptides regions in AGR2 protein (using Disprot) are shown in Figure 4(B). The assay [Fig. 4(A)] included a pre-incubation stage where AGR2 oligomers were incubated with overlapping peptides derived from the open reading frame of AGR2 [Fig. 4(C)]. These overlapping AGR2-derived peptides (numbered 1–5) are from the N-terminal leader sequence, disordered region, and the dimeric interface [Fig. 4(C)]. Peptides 2 and 3 derived from the disordered region stabilize the AGR2 oligomer *in trans*, with peptide 3 inducing the most stabilized oligomer [Fig. 4(D)]. There is a dose-dependence in oligomer stabilization with a concentration of peptide as little as 3 μM increasing stability [Fig. 4(D)]. As controls, peptides overlapping the dimer interface do not stabilize the oligomer. Rather, these peptides 4 and 5 reduce the fluorescence [Fig. 4(E)] with peptide 5 containing the interface sequence 60-EALYK-64 inducing the greatest degree of oligomer de-stabilization. The ability of peptide 4 to similarly de-stabilize the oligomer, although to a lesser extent than peptide 5, suggests other motifs in AGR2 can be targeted to affect the dimer stability.

The common amino acid sequences of peptides 4 and 5 that attenuate oligomer stability include 51-QLIWT-55 [Fig. 5(A)]. The more bioactive peptide 5 has the additional 56-QTYEEALYK-64 comprising the dimer interface [Fig. 5(A)], which suggests why peptide 5 is more active at disrupting the AGR2 oligomer using the ^{2S}MTA. We created a set of AGR2 mutants to determine whether the ^{2S}MTA can be used on full-length mutants to observe changes in oligomer stability. We chose to mutate Y58, since it protrudes into the adjacent internal β-sheet and might be required for stabilizing the dimer interface [Fig. 5(C)]. We also mutated E59 as it protrudes outward into solvent [Fig. 5(C)], whilst E60 forms a salt bridge with K64 across the adjacent monomer forming an important feature of the dimer interface [Fig. 5(A)]. The mutant forms of AGR2 created to determine how these affect oligomer stability, included Y58A, E59A, E60A, and Δ45 (deletion of amino acids 21–45) [Fig. 5(B)]. A titration of AGR2-E59A revealed its' oligomer stability is similar to wt-AGR2 [Fig. 5(B)]. By contrast, AGR2-E60A was in an apparent monomeric state, as judged by the absence of a signal in the oligomer assay [Fig. 5(B)]. The AGR2-Y58A was attenuated as an

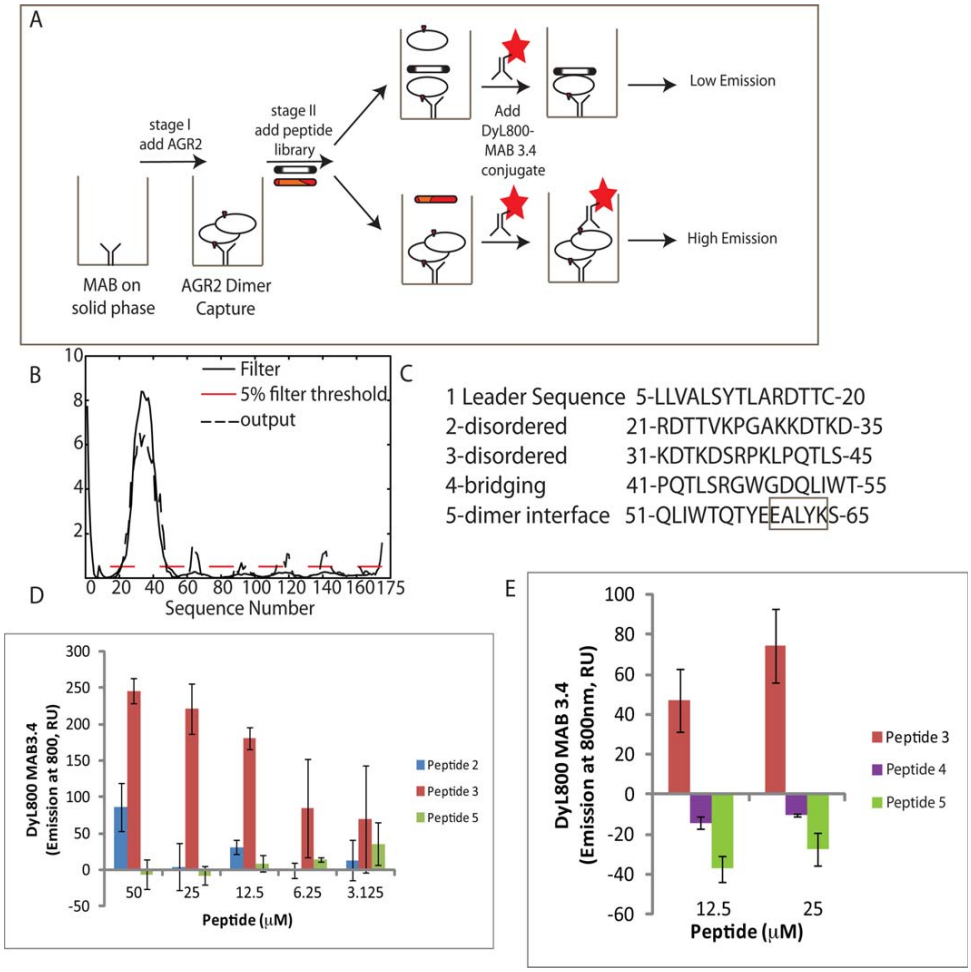


Figure 4. Effects of self-peptides on the extent of AGR2 multimerization. (A) Staging the effects of self-peptides on the oligomerization of AGR2. (B) Plot of the predicted disorder of AGR2 as a function of the amino acid sequence (Disprot). (C). Overlapping peptide sequences from the disordered and dimeric region of AGR2. (D) Effects of the "disordered" peptides 2 and 3 on AGR2 oligomerization. The ^{2S}MTA was set up as in Figure 3(E) except that AGR2 protein (5.5 pmoles) was preincubated with the overlapping peptides 2 and 3 at the indicated concentrations from 3 to 50 μ*M* for 1 h. After washing the DyL800-MAB3.4 probe was added for 1 h and the extent of oligomerization was measured as a function of emission at 800 nm. The data are depicted as DyL800-MAB binding (emission at 800 nm) as a function of peptide 2 and 3 concentration with subtraction of the inert peptide control. (E). Effects of the dimer-interface peptides 4 and 5 on AGR2 oligomerization. AGR2 protein (5.5 pmoles) was preincubated with the overlapping peptides 3, 4, and 5 at the indicated concentrations for 1 h. After washing the DyL800-MAB3.4 probe was added for 1 h and the extent of oligomerization was measured as a function of emission at 800 nm. The data are depicted as DyL800-MAB binding (emission at 800 nm) as a function of peptide 3, 4, and 5 concentration after subtraction of the inert peptide control background. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

oligomer with a signal in the ^{2S}MTA only seen at the highest concentration of the protein [Fig. 5(B)]. By contrast to these loss-of-function mutants, Δ45-AGR2 was more stable as a dimer [Fig. 5(B)]. The enhanced stability of the oligomer by deletion of the disordered region is consistent with prior research showing that deletion of amino acids 21–40 increases dimer stability.[23]

Mapping the crosslinked AGR2 dimer

We sought to develop an independent assay to validate the stability of the AGR2 dimer through the use of a DSS crosslinker that can trap a covalent dimeric species [Fig. 6(A)]. Mapping of this crosslink using mass spectrometry identified K95 as the dominant cross-link [Fig. 6(B,C)]. Mutagenesis of K95 to create AGR2-K95R confirmed that this is the dominant cross-link site [Fig. 6(D)]. An image of the position of K95 relative to the dimer interface is depicted in Figure 6(E). We used this assay to determine whether the AGR2 mutant panel exhibited altered dimerization (Fig. 7). A titration of DSS into reactions containing wt-AGR2, AGR2-E60A, and Δ45-AGR2 demonstrated that AGR2-E60 is attenuated in

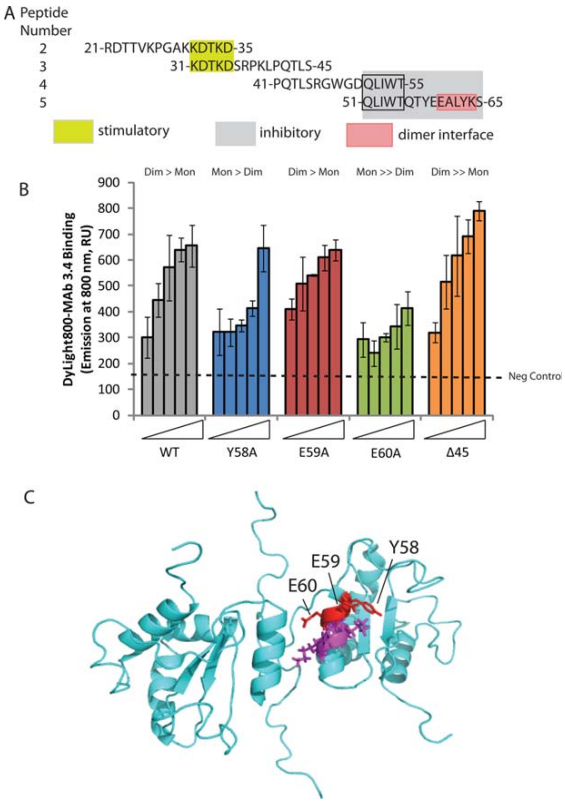


Figure 5. Effects of single point mutations in AGR2 dimer interface on multimerization using ^{2S}MTA. (A) Sequence of bioactive peptides (from Fig. 4) derived from the disordered and dimerization region of AGR2 protein. Highlighted are minimal motifs predicted to drive stimulation (KDTKD from peptides 2 and 3) and inhibition (QLIWT from peptides 4 and 5). The dimer interface (from Ref. 23) in peptide 5 is highlighted as EALYK. (B). Oligomer stability in AGR2 mutant panel. Increasing amounts of wild-type and mutant AGR2 proteins (from left, 0, 0.69, 1.38, 2.75, and 5.5 pmoles) were captured using MAB3.4 as indicated in Figure 3(E). Following the addition of DyL800-MAB3.4, the extent of oligomerization was quantified by emission at 800 nm and the data is presented as an average from triplicate titrations. No protein background signal is represented on the far left (157 U). The ratio of dimer to monomer is highlighted above each series: wt: dimmer > monomer; Y58A: monomer > dimer; E59: as wild-type; E60A: monomeric; and Δ45: more stable dimeric over monomeric. (C). Summary of the positions of Y58, E59, and E60 across the dimer interface region. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

dimer formation [Fig. 7(B)]. In addition, Δ45-AGR2 exists in a more stable, spontaneous dimer in the absence of cross link [Fig. 7(C), lane 1], suggesting that deletion of the N-terminus can allow it to form a stable dimer in the presence of the 0.1% SDS buffer. These data together are consistent with the quantitative ^{2S}MTA that demonstrated Δ45-AGR2 is a more stable dimer whilst AGR2-E60A is predominantly monomeric.

Evaluating the specific activity of the AGR2 oligomer

There is no known biochemical function for the AGR2 dimer. Having created a series of gain-of-function or loss of function AGR2 mutants, we evaluated the role of oligomerization in molecular interaction of AGR2 with two of its most well-characterized substrates: peptides containing the [T/S]-x-I-[YF]-[Y/F] consensus motif[20] and the AAA+ chaperone protein, Reptin.[26] The former protein-interaction presumably represents docking sites in its client proteins (manuscript in preparation). The latter protein interaction is with the AAA+ protein Reptin, which itself functions as a molecular machine able to assemble multiprotein complexes. The *in vivo* impact of AGR2 protein binding on client proteins containing the [T/S]-x-I-[Y/F]-[Y/F] motif and on Reptin complex assembly is currently undefined. Nevertheless, we wanted to determine whether AGR2 binding to either of these two targets was influenced by its' dimer-monomer state. A titration of wt-AGR2, AGR2-E60A, and Δ45-AGR2 indicates that all three exhibited relatively similar activity in binding to this peptide substrate [Fig. 8(A)]. By contrast, the Reptin substrate is substantially stabilized using the gain-of-function more stable dimer mutant Δ45-AGR2, whilst AGR2-E60A is attenuated [Fig. 8(B)]. These data provide the first biochemical evidence that dimerization of AGR2 protein can play a role in its affinity for a client protein. As there are over one dozen binding proteins identified for AGR2 protein using yeast-two hybrid,[21] it will be interesting to see whether this panel of AGR2 oligomerization mutants can be used to access the role of the AGR2 dimer in these protein-protein interactions.

Discussion

Developing assays that measure oligomerization and how allosteric effects can be measured in parallel will provide important tools in the emerging field of drugging protein-protein interactions. We evaluate here an approach to measure oligomerization and allostery using Anterior Gradient-2. The mechanism whereby AGR2 mediates its' functions is thought to include its ability to chaperone proteins in the ER[27] and to mediate induction of oncogenic signals like EGF or p53 tumor suppressor silencing.[28,29] The majority of AGR2 protein interactions have been identified using yeast-two hybrid,[21] with the only well-validated protein being the AAA+ pro-metastatic protein Reptin.[26] Characterizing these various protein–protein interactions will likely shed light on its mechanism of action. Toward this aim, we focused on characterizing the AGR2 oligomer (dimer) as a model, since our initial biochemical characterizations demonstrated that its apparent mass was consistent with a homodimeric quaternary structure (Fig. 1). This is consistent with a recent report showing that recombinant AGR2

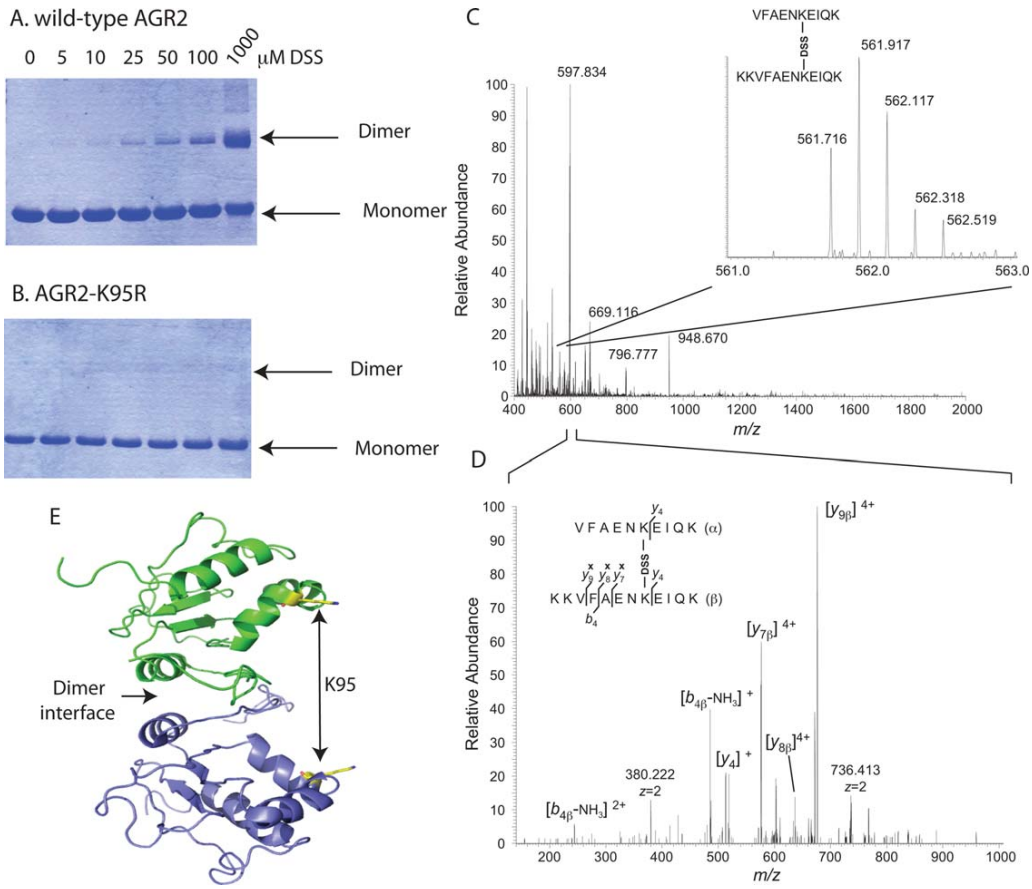


Figure 6. Mapping the sites of DSS dimer crosslinking to Lysine-95. (A) Crosslinking of AGR2 into a homodimeric species using the covalent crosslinker DSS. (B) Absence of crosslinking of AGR2-K95R mutation into a homodimeric species using the covalent crosslinker DSS, based on the K95-K95 crosslinks identified in C and D. (C). Mass spectrum of interprotein crosslinked peptides VFAEN**K**EIQK and KKVFAEN**K**EIQK (the bold K corresponds to the DSS-linked amino-acid), at $[M+5H]^{5+} = 561.716$ *m/z* units. (D) The lower panel shows the annotated tandem mass spectrum after coalitional-induced dissociation and subsequent Orbitrap mass analysis. Crosslinked peptides were identified as indicated in the materials and methods. (E) Cartoon of the interdomain orientation of the K95 amino acid residues with respect to the dimeric interface. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

($AGR2^{21–175}$ or $AGR2^{41–175}$) lacking the N-terminal leader sequence or disordered region, respectively, can form homodimers.[23] Showing that AGR2 can exist as a dimer *in vitro*, we then wanted to determine whether the dimeric structure was important for two of the most well-characterized interactions; (i) the [T/]-x-I-[Y/F]-[Y/F] pentapeptide consensus binding site; and (ii) the AAA+ protein Reptin. Toward this aim, we needed to create a set of mutants in AGR2 with altered oligomeric state and we needed robust assays to define changes in its dimeric structure as a function of mutation.

The assay designed to measure oligomer stability was a two-site microtiter assay in which an oligomeric protein captured in solid phase can be detected in small volume liquid phase using the same monoclonal antibody; for example a "sandwich assay" (^{2S}MTA). In this assay it is not possible for a monomer to give a signal. The ^{2S}MTA detected attenuation of AGR2 oligomerization using self-peptides derived from the dimer interface (60-EALYK-64). Amino acids outwith this motif also attenuated oligomer stability (e.g., peptide 4; Fig. 4) suggesting that the dimer motif requires the stability of the adjacent peptide chain. In addition, we also surprisingly identified peptide motifs that can stimulate AGR2 oligomerization *in trans*. These peptides are derived from the intrinsically disordered region in the N-terminal domain of AGR2 (Fig. 4) and their ability to act *in trans* as positive effectors of oligomerization suggests they disrupt the function of the natural N-terminal sequences *in cis*. There are interesting paradigms where peptides and mimetics can allosterically stimulate protein–protein interactions[30,31] and it will be interesting to determine whether these AGR2 stimulatory peptides can be used as tools to manipulate the activities of the protein. Using ^{2S}MTA as a rapid screen for oligomerization capability, we confirmed that AGR2-E60A was

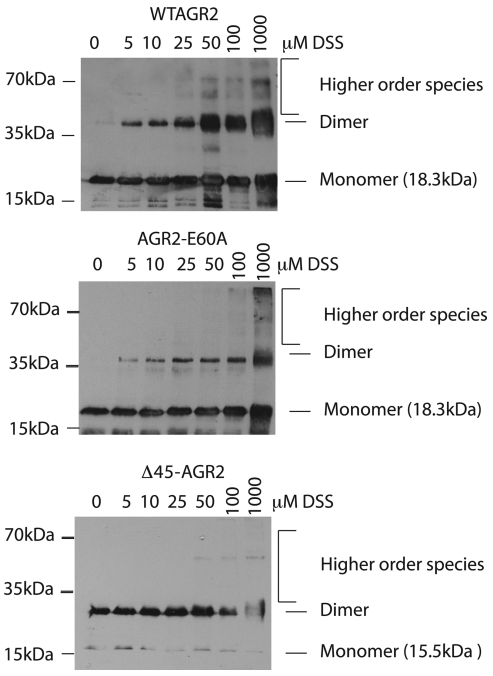


Figure 7. Effects of DSS on the dimerization of AGR2 mutants. The indicated proteins; (A) wt-AGR2, (B), AGR2-E60A, or (C) Δ45-AGR2, were incubated with increasing concentrations of DSS and separated by electrophoresis on an SDS gel (containing 0.1% SDS). AGR2 was detected using MAB3.4 in an immunoblot. The positions of monomeric and dimeric proteins are highlighted.

largely monomeric (Fig. 5) and Δ45-AGR2 had an enhanced oligomerization capacity (Fig. 5). This was confirmed using a DSS crosslinking assay (Fig. 7) that was mapped to the K95 residue using mass spectrometry (Fig. 6).

Having developed a set of well-characterized oligomerization mutants, the effect of subunit structure on biochemical activity was finally evaluated. The AGR2 monomeric mutant E60A bound as stably as wild-type AGR2 or the enhanced dimer mutant Δ45-AGR2 to the [T/S]-x-I-[Y/F]-[Y/F] pentapeptide motif. This data indicates that if the TxIYY docking site represents a cellular binding interface for AGR2 client proteins, then the monomeric form of AGR2 has the potential to be equally active. By contrast, the binding of AGR2 to Reptin was stabilized by creating an enhanced dimer (Δ45-AGR2) and attenuated using the monomeric E60A mutant AGR2. Thus the determinant in AGR2 that interacts with Reptin (primarily within amino acids 104–111;[26]) exploit its' dimeric subunit structure. Solving the interface of AGR2-Reptin might facilitate developing biologics that disrupt this protein interaction to determine how it might contribute to cancer cell growth.

In conclusion, we have developed a microtiter-based assay that measures in small volumes the oligomerization state of AGR2 protein. Such an assay was used to identify ligands like self-peptides that can allosterically regulate multimer stability thus forming a proof-of-concept assay for screening for cellular proteins or drug-leads that might regulate of AGR2 dimer stability *in vivo*. The use of microtiter assays like the "sandwich"-type assay (^{2S}MTA) that measures ligand-free oligomerization provides a robust small volume methodology that can be adapted to screen for ligands that regulate the stability of oligomeric proteins implicated in human disease.

Experimental

Purification of recombinant AGR2 protein and analytical gel filtration size-exclusion chromatography

AGR2$^{21–175}$ was cloned into pEHISTEV and transformed into BL21 (DE3).[20,32] Bacteria were grown to OD_{600} of 0.6 at 37°C and 1 m*M* IPTG was added for a further 3 h, followed by harvesting and resuspending in 30 mL of lysis buffer (50 m*M* NaH_2PO_4 pH 7.5, 400 m*M* NaCl, 10 m*M* Imidazole) on ice. Cells were subjected to high-pressure (25k PSI) followed by centrifugation at 40,000*g* for 50 min to isolate the soluble fraction. The supernatant was applied to a Ni-charged HisTrap FF 5-mL column (GE Life Sciences). The immobilized protein was subject to a step-elution with 30 m*M* then 300 m*M* Imidazole. The eluted protein was pooled and cleaved with recombinant tobacco etch virus protease overnight at 20°C. The resultant solution was desalted into protein buffer (50 m*M* Tris pH 7.5, 200 m*M* NaCl) with a HiPrep desalt 26/10 column (GE Life Sciences) before reloading onto the IMAC to remove protease and uncleaved protein. Purified protein was analyzed by gel-filtration size exclusion chromatography on an analytical Superdex PC 75 3.2/30" column (GE Healthcare) at titrated concentrations ranging from 2.5 mg mL^{-1} to 5 μg mL^{-1} (injection volume of 20 μL). Detection was by absorption at 214 nm at a flow rate of 0.5 mL min^{-1}. The column was calibrated using protein standards (GE Healthcare).

Monoclonal antibody purification and DyLight800 labeling. Monoclonal antibody MAB3.4[25] was affinity purified on a Protein G column by elution using a buffer containing 0.1 M Glycine (pH 2.5) and then neutralized with 1/10th volume of a 1 M Tris HCl (pH 8.8) buffer. The IgG was dialyzed against PBS and was conjugated to DyLight800 fluorophore (excitation at 770 nm and emission at 794 nm) using DyLight800 Microscale Antibody Labeling Kit (53063; Thermo Scientific). Briefly, 50 μL of MAB3.4 (1.3 mg mL^{-1}) was diluted to 100 μL final volume with 0.67*M* Borate buffer and the protein as added a vial of DyLight800 reagent to initiate coupling and the reaction is continued in the dark for 60 min at room temperature. The labeled protein

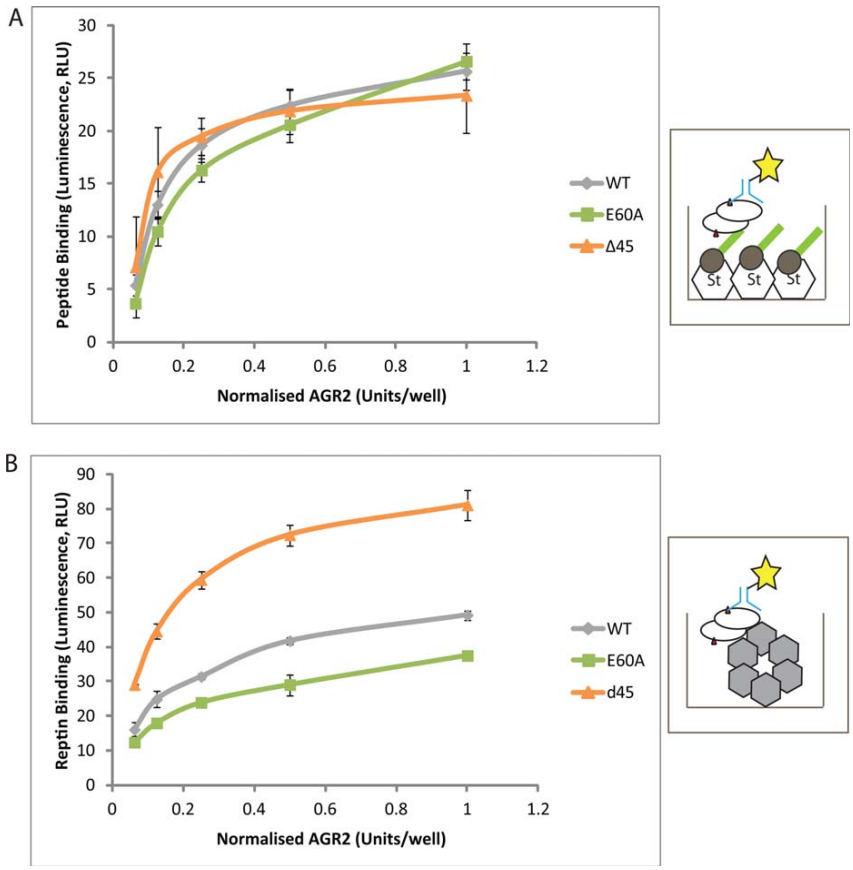


Figure 8. The effects of AGR2 oligomerization on its protein-interaction activity. The indicated AGR2 proteins (wt-AGR2, AGR2-E60A, or Δ45-AGR2) were incubated in reactions containing (A) biotinylated TxIYY containing peptide and (B) Reptin. AGR2 bound to either peptide or Reptin was detected using MAB3.4 and quantified using anti-secondary IgG coupled to peroxidase and chemiluminescence. The data are depicted as AGR2 binding activity (in RLU) as a function of increasing AGR2 protein concentration (in triplicate and presented as an average). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

was added to a spin column in microfuge format and the labeled antibody was separated from the label by collecting the eluate after centrifugation.

Measuring AGR2 oligomerization using a two-site sandwich microtiter assay (^{2S}MTA). Immunoglobulins (100 ng/50 ml per well) were adsorbed onto 96-well Costar opaque (white) wells, diluted in 100 m*M* $NaHCO_3$ pH 8.6 and incubated at 4°C, with agitation, for 14–16 h. Wells were washed 3–5× with PBS-0.05% Tween-20, prior to blocking with 200 μL of 3% bovine serum albumin in PBS-0.05% Tween-20 per well for 1 h at room temperature. Protein was titrated in triplicate to the desired concentration in blocking buffer to 50 μL per well and incubated on the plate for 1 h at room temperature and washed 3–5× with PBS-0.05% Tween-20 prior to antibody detection. Primary antibody was diluted 1:1000 (1 μg mL^{-1}) and 50-μL per well incubated at room temperature for 1 h, plates were then washed 3–5× with PBS-0.05% Tween-20. For ECL, 50 μL secondary antibody (rabbit anti-mouse, 1:1000 in 3% BSA-PBST) was incubated for 1 h, prior to a final 3–5× washes with PBS-0.05% Tween-20 and relevant detection. To detect dimerization: (i) enhanced chemiluminescence was carried out using white 96-well plates and secondary antibody conjugated to horseradish peroxidase followed by detection on a Fluoroskan Accent FL (ii) alternatively, a 800 nm near-IR label (DyLight 800, Thermo) was used to label the anti-AGR2 MAB3.4 and, using black walled, clear bottomed 96-well microtiter plates (Costar), antibody binding was quantified using Licor Odyssey.

Oligomer disruption assays using ^{2S}MTA

The capture MAB (100 ng per well of MAB3.4) was diluted in 50 μL of $NaHCO_3$ (pH 9.0) buffer and adsorbed on to the black walled, clear bottomed 96-well microtiter plate overnight at 4°C. Plates were then washed 3–5× with PBS-0.05% Tween-20, and blocked with 3% BSA in PBS-0.05% Tween-20 for 1 h. Concurrently, 100 ng (up to 5.5 pmoles) of

purified AGR2 protein was diluted to 50 μL of protein buffer, and incubated with the indicate amounts of biotinylated overlapping peptide (in triplicate) for 1 h. The blocking solution was removed from the plate and replaced by peptide–protein solution and incubated for 1 h. Nearly 150 ng of DyLight800-labeled MAB3.4-labeled was then added to each well and allowed to bind for 1 h. Unbound or weakly bound antibody was then removed by 5× washes with PBS-0.05% Tween-20 before directly detected on the Licor Odyssey with 750 nm laser excitation and emission at 800 nm. Images were analyzed and quantified with the Odyssey Sa software.

Site *directed mutagenesis of AGR2$^{21–175}$-pEHISTEV to produce Y58A, E59A, and E60A mutants and cloning for Δ45 mutant.* The indicated primers (Y58A, E59A, and E60A) were used for mutagenesis using GENEART mutagenesis kit (*Invitrogen*).

Mutation	Forward primer	Reverse primer
Y58A	ATCTGGACTCAGAC AGCTGAAGAAGCT CTATAT	ATATAGAGCTTCTTC AGCTGTCTGAGTC CAGAT
E59A	TGGACTCAGACATA TGCAGAAGCTCTA TATAAA	TTTATATAGAGCTTC TGCATATGTCTGA GTCCA
E60A	ACTCAGACATATGA AGCAGCTCTATAT AAATCC	GGATTTATATAGAG CTGCTTCATATGT CTGAGT

The Δ45 mutant was amplified by PCR using forward (CCATGGCTATGAGAGGTTGGGGTGACCAAC) and reverse (CTCGAGTTACAATTCAGTCTTCAGCAACTT). The PCR products were cloned into pEHISTEV vector digested with *Nco*I and *Xho*I restriction enzymes to generate the final vectors.

Measuring mutant AGR2 protein oligomerization status using ^{2S}MTA. About 500 mL pellets of bacteria induced to express wild-type and mutant AGR2 were resuspended in 10 mL lysis buffer and subject to 3 × 10 s sonication on ice. Protein mutants included Y58A, E59A, E60A, and Δ45-AGR2. The lysed bacteria were centrifuged to isolate the soluble fraction. About 10 μL fractions of each mutant crude lysate was separated by SDS-PAGE, transferred to nitrocellulose and probed with the DyLight800-labeled MAB3.4. Using the Licor Odyssey system an immunoblot was used to quantify and normalize AGR2 protein concentrations (as in Fig. 5). Nearly 100 ng of MAB3.4 antibody was adsorbed onto a clear-bottomed 96-well plate and the plate was washed and blocked as described previously. A titration of the normalized AGR2 protein (in triplicate) was applied to the plate in the mobile phase before washing 5× PBS-0.05% Tween-20 and quantification using the DyLight800-labeled MAB3.4.

Measuring the AGR2 oligomerization status using the K95-dependent DSS cross-linking assay. For protein crosslinking, AGR2 protein was exposed to DSS (Pierce 21555) and separated by SDS gel electrophoresis. The DSS crosslinked protein was dissolved in 50 m*M* ammonium bicarbonate and 6 M urea, reduced with 1 μL 50 m*M* DTE (dithioerythritol) for 60 min at 37°C, and alkylated by adding 1 μL of 400 m*M* iodoacetamide for 60 min at room temperature. The sample was diluted three times and digested overnight at 37°C with 0.1 μg of sequencing grade trypsin (Promega, Madison, WI). The remaining methodology for identification of crosslinked peptides has been described,[33–35] including the use of an LTQ Orbitrap velos mass spectrometer (Thermo Electron; San Jose, CA) equipped with a NanoAcquity system (Waters). Mutagenic primers for K95R mutant generation were as follows: K95R forward: 5′-gtgtttgctgaaaatagagaaatccagaaattg, and reverse: 5′-caatttctggatttctctattttcagcaaacac.

Measuring AGR2 protein–protein interactions. AGR2-Reptin. Human Reptin expressed in BL21 AI (Invitrogen) *E. coli* was purified as described.[26] To measure the binding of AGR2 (and mutants), purified Reptin (100 ng per well) was adsorbed to the well in carbonate buffer (0.1 M pH 9.0) overnight. Following blocking with PBS containing 0.05% Tween-20 and 3% BSA, the indicated AGR2 proteins were titrated in triplicate and incubated for 1 h at room temperature followed by detection with an AGR2 monoclonal antibody (Abnova) and HRP conjugated rabbit anti-mouse secondary antibody (Dako).

AGR2-TxIYY peptide. The TxIYY containing peptide (Biotin-SGSG-HLPTTIYYGPPG;[20]) was captured on streptavidin-coated wells and incubated for 1h, then plates were washed 5× with PBS-0.05% Tween-20 and blocked. AGR2 protein and the indicated mutants were titrated in triplicate onto the plates and incubated for 1 h at room temperature followed by detection using the AGR2 monoclonal antibody and HRP conjugated rabbit anti-mouse secondary antibody.

References

1. Morelli X, Hupp T (2012) Searching for the Holy Grail; protein–protein interaction analysis and modulation. EMBO Rep 13:877–879.
2. Thangudu R, Bryant SH, Panchenko R, Madej T (2012) Modulating protein–protein interactions with small molecules: the importance of binding hotspots. J Mol Biol 415:443–453.
3. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. J Mol Biol 293:321–331.
4. Jubb H, Higueruelo AP, Winter A, Blundell TL (2012) Structural biology and drug discovery for protein–protein interactions. Trends Pharmacol Sci 33:241–248.

5. Winter A, Higueruelo AP, Marsh M, Sigurdardottir A, Pitt WR, Blundell TL (2012) Biophysical and computational fragment-based approaches to targeting protein–protein interactions: applications in structure-guided drug discovery. Q Rev Biophys 45:383–426.
6. Smith MC, Gestwicki JE (2012) Features of protein–protein interactions that translate into potent inhibitors: topology, surface area and affinity. Expert Rev Mol Med 14:e16.
7. Basse MJ, Betzi S, Bourgeas R, Bouzidi S, Chetrit B, Hamon V, Morelli X, Roche P (2013) 2P2Idb: a structural database dedicated to orthosteric modulation of protein–protein interactions. Nucleic Acids Res 41: D824–D827.
8. Anastasiou D, Yu Y, Israelsen WJ, Jiang JK, Boxer MB, Hong BS, Tempel W, Dimov S, Shen M, Jha A, Yang H, Mattaini KR, Metallo CM, Fiske BP, Courtney KD, Malstrom S, Khan TM, Kung C, Skoumbourdis AP, Veith H, Southall N, Walsh MJ, Brimacombe KR, Leister W, Lunt SY, Johnson ZR, Yen KE, Kunii K, Davidson SM, Christofk HR, Austin CP, Inglese J, Harris MH, Asara JM, Stephanopoulos G, Salituro FG, Jin S, Dang L, Auld DS, Park HW, Cantley LC, Thomas CJ, Vander Heiden MJ (2012) Pyruvate kinase M2 activators promote tetramer formation and suppress tumorigenesis. Nat Chem Biol 8:839–847.
9. Morgan HP, O'Reilly FJ, Wear MA, O'Neill JR, Fothergill-Gilmore L, Hupp T, Walkinshaw MD (2013) M2 pyruvate kinase provides a mechanism for nutrient sensing and regulation of cell proliferation. Proc Natl Acad Sci USA 110:5881–5886.
10. Jaffe EK, Lawrence SH (2012) Allostery and the dynamic oligomerization of porphobilinogen synthase. Arch Biochem Biophys 519:144–153.
11. Lawrence, SH, Ramirez UD, Selwood T, Stith L, Jaffe EK (2009) Allosteric inhibition of human porphobilinogen synthase. J Biol Chem 284:35807–35817.
12. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 37:215–246.
13. Tompa P (2012) On the supertertiary structure of proteins. Nat Chem Biol 8:597–600.
14. Kovacs D, Szabo B, Pancsa R, Tompa P (2012) Intrinsically disordered proteins undergo and assist folding transitions in the proteome. Arch Biochem Biophys 531:80–89.
15. Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, Dosztanyi Z, Uversky VN, Obradovic Z, Kurgan L, Dunker AK, Gough JD (2013)(2)P(2): database of disordered protein predictions. Nucleic Acids Res 41:D508–D516.
16. Poyurovsky MV, Priest C, Kentsis A, Borden KL, Pan ZQ, Pavletich N, Prives C (2007) The Mdm2 RING domain C-terminus is required for supramolecular assembly and ubiquitin ligase activity. EMBO J 26:90–101.
17. Kostic M, Matt T, Martinez-Yamout MA, Dyson HJ, Wright PE (2006) Solution structure of the Hdm2 C2H2C4 RING, a domain critical for ubiquitination of p53. J Mol Biol 363:433–450.
18. Maslon MM, Hupp TR (2010) Drug discovery and mutant p53. Trends Cell Biol 20:542–555.
19. Robson AF, Hupp TR, Lickiss F, Ball KL, Faulds K, Graham D (2012) Nanosensing protein allostery using a bivalent mouse double minute two (MDM2) assay. Proc Natl Acad Sci USA 109:8073–8078.
20. Murray E, McKenna EO, Burch LR, Dillon J, Langridge-Smith P, Kolch W, Pitt, A, Hupp TR (2007) Microarray-formatted clinical biomarker assay development using peptide aptamers to anterior gradient-2. Biochemistry 46:13742–13751.
21. Chevet E, Fessart D, Delom F, Mulot A, Vojtesek B, Hrstka R, Murray E, Gray, T, Hupp T (2013) Emerging roles for the pro-oncogenic anterior gradient-2 in cancer development. Oncogene 32:2499–2509
22. Fourtouna A, Murray E, Nicholson J, Maslon MM, Pang L, Dryden DTF, Hupp TR (2009) The anterior gradient-2 pathway as a model for developing peptide-aptamer anti-cancer drug leads that stimulate p53 function. Curr Chem Biol 3:124–137.
23. Patel P, Clarke, C, Barraclough DL, Jowitt TA, Rudland PS, Barraclough R, Lian, LY (2012) Metastasis-promoting anterior gradient 2 protein has a dimeric thioredoxin fold structure and a role in cell adhesion. J Mol Biol 425:929–943.
24. Hengel SM, Murray E, Langdon S, Hayward L, O'Donoghue J, Panchaud A, Hupp T, Goodlett DR (2011) Data-independent proteomic screen identifies novel tamoxifen agonist that mediates drug resistance. J Proteome Res 10:4567–4578.
25. Gray TA, MacLaine NJ, Michie CO, Bouchalova P, Murray E, Howie J, Hrstka R, Maslon MM, Nenutil R, Vojtesek B, Langdon S, Hayward L, Gourley C, Hupp TR (2012) Anterior gradient-3: a novel biomarker for ovarian cancer that mediates cisplatin resistance in xenograft models. J Immunol Methods 378:20–32.
26. Maslon MM, Hrstka R, Vojtesek B, Hupp TR (2010) A divergent substrate-binding loop within the pro-oncogenic protein anterior gradient-2 forms a docking site for Reptin. J Mol Biol 404:418–438.
27. Higa A, Mulot A, Delom F, Bouchecareilh M, Nguyen DT, Boismenu D, Wise MJ, Chevet E (2011) Role of pro-oncogenic protein disulfide isomerase (PDI) family member anterior gradient 2 (AGR2) in the control of endoplasmic reticulum homeostasis. J Biol Chem 286: 44855–44868.
28. Gupta A, Dong A, Lowe AW (2012) AGR2 gene function requires a unique endoplasmic reticulum localization motif. J Biol Chem 287:4773–4782.
29. Pohler E, Craig AL, Cotton J, Lawrie, L, Dillon JF, Ross P, Kernohan N, Hupp TR (2004) The Barrett's antigen anterior gradient-2 silences the p53 transcriptional response to DNA damage. Mol Cell Proteomics 3:534–547.
30. Bouchecareilh M, Higa A, Fribourg S, Moenner M, Chevet E (2011) Peptides derived from the bifunctional kinase/RNase enzyme IRE1alpha modulate IRE1alpha activity and protect cells from endoplasmic reticulum stress. FASEB J 25:3115–3129.
31. Nicholson J, Neelagandan K, Huart AS, Ball K, Molloy MP, Hupp TR (2012) An iTRAQ proteomics screen reveals the effects of the MDM2 binding ligand Nutlin-3 on cellular proteostasis. J Proteome Res 11: 5464–5478.
32. Liu H, Naismith JH (2009) A simple and efficient expression and purification system using two newly constructed vectors. Protein Expr Purif 63:102–111.
33. Singh P, Shaffer SA, Scherl A, Holman C, Pfuetzner RA, Larson Freeman TJ, Miller SI, Hernandez P, Appel RD, Goodlett DR (2008) Characterization of protein cross-links via mass spectrometry and an open-modification search strategy. Anal Chem 80:8799–8806.
34. Tsutsui Y, Wintrode PL (2007) Hydrogen/deuterium exchange-mass spectrometry: a powerful tool for probing protein structure, dynamics and interactions. Curr Med Chem 14:2344–2358.
35. Marin M, Thallmair V, Ott T (2012) The intrinsically disordered N-terminal region of AtREM1.3 remorin protein mediates protein–protein interactions. J Biol Chem 287:39982–39991.

Chemical Science

EDGE ARTICLE

View Article Online
View Journal | View Issue

CrossMark click for updates

Cite this: *Chem. Sci.*, 2015, **6**, 3109

Discovery of a novel ligand that modulates the protein–protein interactions of the AAA+ superfamily oncoprotein reptin†

Alan R. Healy,[a] Douglas R. Houston,*[b] Lucy Remnant,[c] Anne-Sophie Huart,[c] Veronika Brychtova,[d] Magda M. Maslon,[c] Olivia Meers,[c] Petr Muller,[d] Adam Krejci,[d] Elizabeth A. Blackburn,[b] Borek Vojtesek,[d] Lenka Hernychova,[d] Malcolm D. Walkinshaw,[b] Nicholas J. Westwood*[a] and Ted R. Hupp*[c]

Developing approaches to discover protein–protein interactions (PPIs) remains a fundamental challenge. A chemical biology platform is applied here to identify novel PPIs for the AAA+ superfamily oncoprotein reptin. An *in silico* screen coupled with chemical optimization provided Liddean, a nucleotide-mimetic which modulates reptin's oligomerization status, protein-binding activity and global conformation. Combinatorial peptide phage library screening of Liddean-bound reptin with next generation sequencing identified interaction motifs including a novel reptin docking site on the p53 tumor suppressor protein. Proximity ligation assays demonstrated that endogenous reptin forms a predominantly cytoplasmic complex with its paralog pontin in cancer cells and Liddean promotes a shift of this complex to the nucleus. An emerging view of PPIs in higher eukaryotes is that they occur through a striking diversity of linear peptide motifs. The discovery of a compound that alters reptin's protein interaction landscape potentially leads to novel avenues for therapeutic development.

Received 15th December 2014
Accepted 20th March 2015

DOI: 10.1039/c4sc03885a

www.rsc.org/chemicalscience

Open Access Article. Published on 20 March 2015. Downloaded on 29/06/2017 13:01:39.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Introduction

Discovering protein–protein interactions (PPIs) remains a major challenge.[1] However, a detailed understanding of a protein's PPI network is central to novel biomedical applications. Rate-limiting protein nodes need to be identified as they can serve as a focus for novel diagnostic and/or therapeutic advances. The current view is that drugging PPIs remains an untapped landscape in the drug discovery field.

The main approaches used to define the PPIs of a target protein include yeast two-hybrid methods[2] and mass-spectrometry-based sequencing of multi-protein complexes using tagged-bait proteins.[3] Although powerful, their limitations are that they are done outside an authentic cellular context, require artificial tagging of the bait protein and are unable to capture weak, or dynamic interactions. One advance in the study of PPIs is the idea that a large proportion of the polypeptide sequence information in higher eukaryotes is intrinsically disordered thus providing a template for "weak" regulatory, combinatorial and specific PPIs to occur in signal transduction.[4] A second advance is the realization that a number of PPIs occur *via* a linear amino acid motif[1] that provides opportunities for sequence based hotspots to be identified. Developing methods to capture such consensus linear motifs of a target protein would complement the technology currently used to discover PPIs.

The AAA+ (ATPase associated with various cellular activities) superfamily of proteins is present within all kingdoms of life.[5] Members of this family exist as oligomers and form compelling targets in understanding allosteric control of protein function.[6] Reptin and pontin represent two highly conserved members that are now viewed as model systems to define fundamental aspects of AAA+ superfamily function in eukaryotes.[7] Reptin is an important regulator of key cellular functions through a range of PPIs.[8–19] The different oligomeric forms of reptin and its ability to form a range of complexes with different compositions could underpin its functional diversity.[20–22] The composition of these oligomeric complexes must be tightly regulated and this has been linked to reptin's bound ligand ATP/ADP.[23,24] Development of a synthetic ATP/ADP mimetic to probe the intrinsic oligomerization properties of reptin and its ability to form diverse PPIs could provide insight into the regulation and function of reptin and the wider AAA+ family.

[a]*School of Chemistry & Biomedical Sciences Research Complex, University of St Andrews & EaStCHEM, North Haugh, St Andrews, KY16 9ST, UK. E-mail: njw3@st-andrews.ac.uk*

[b]*Centre for Chemical Biology, University of Edinburgh, EH9 3JG, UK. E-mail: DouglasR.Houston@ed.ac.uk*

[c]*Edinburgh Cancer Research Centre, Cell Signalling Unit, University of Edinburgh, EH4 2XR, UK. E-mail: ted.hupp@ed.ac.uk*

[d]*RECAMO, Masaryk Memorial Cancer Institute, 656 53 Brno, Czech Republic*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c4sc03885a

Open Access Article. Published on 20 March 2015. Downloaded on 29/06/2017 13:01:39.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Here, a chemical biology platform is used to deliver a novel tool to dissect the function of reptin. This approach involves; (i) *in silico* screening of virtual libraries to identify novel ATP mimetics; (ii) optimization of a hit as a PPI and oligomerization modifier giving the novel chemical tool, Liddean; (iii) use of combinatorial-peptide libraries and next generation sequencing to identify novel responsive PPIs; and (iv) cell based validation of ligand-activated PPIs using proximity ligation assays. We also demonstrate that our chemical tool Liddean (an ATP-mimetic) can be used to discover and manipulate the PPIs of the AAA+ protein reptin.

Results

In silico screening leading to a prioritized ATP mimetic

An *in silico* screen was used to identify small molecules that might bind at the Walker A site on reptin. The rigid-body

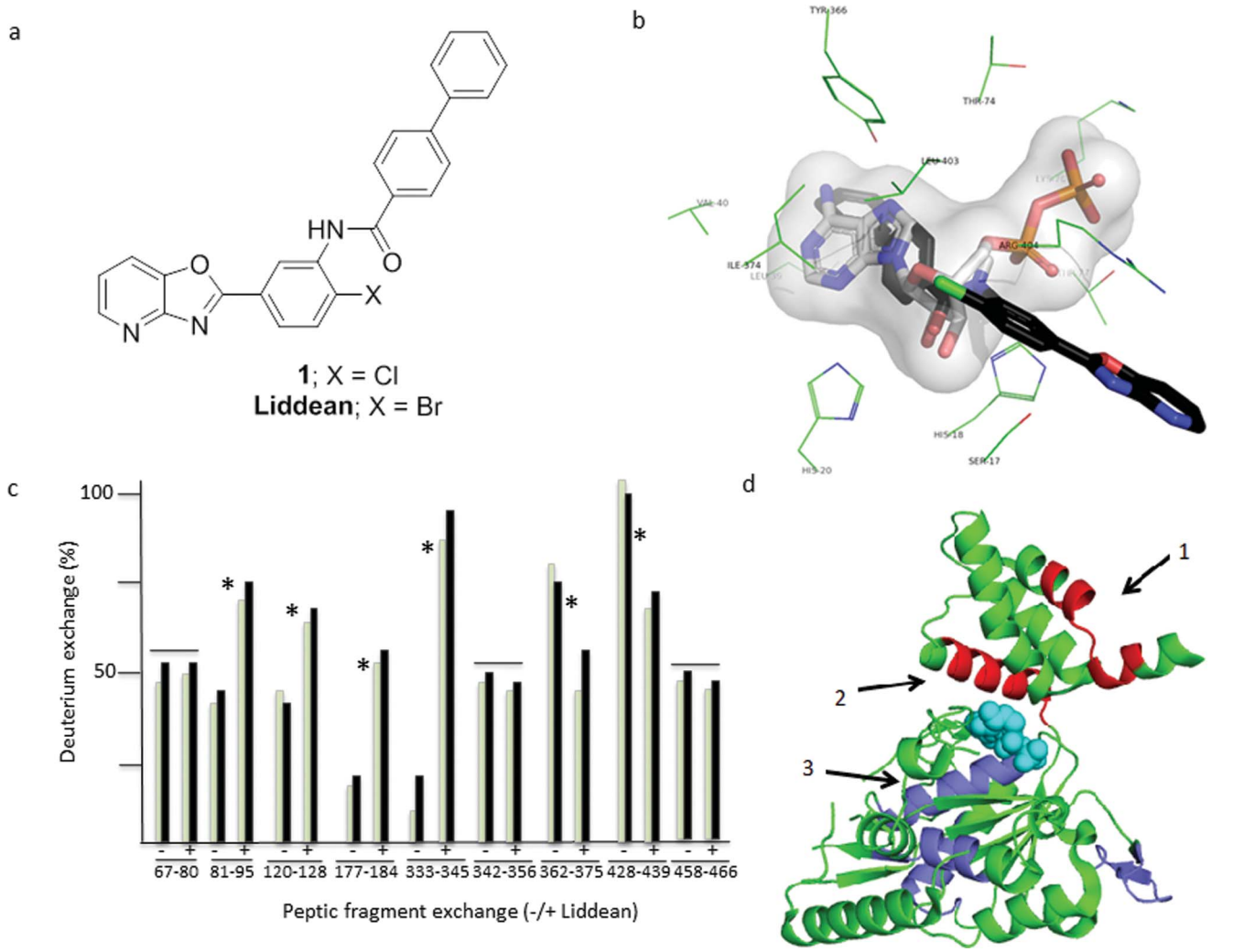


Fig. 1 Identification of small molecules targeting reptin using an *in silico* screening programme. (a) Structure of hit compound **1** was identified through a reptin–AGR2 peptide interaction assay of the top 30 hits from the *in silico* screen (see Fig. S1 and S2†) and Liddean, the most active analog obtained through chemical optimization (see Scheme S1 and Fig. S4 and S5†). (b) The Autodock prediction of the binding mode of **1** is shown as black sticks. The side chains of residues that comprise the active site are shown as green lines and are labeled. The ADP molecule is colored white and shown as sticks. The pocket is also shown as a transparent surface representation. In all cases nitrogen is colored blue, oxygen red, phosphorous orange, and chlorine green. See also Fig. S1 and S3.† (c) Changes in hydrogen–deuterium exchange on reptin peptide motifs after pepsin proteolysis. Ligand free reptin protein (Fig. S7a†) was digested with pepsin after processing in the absence or presence of Liddean. Processing of peptides was performed using HPLC-MS/MS with the instrument operated in a data-dependent mode. All identified peptides are shown as green bars (1 minute) or black bars (five minutes) without or with Liddean (L) and the % change in deuteration as a function of peptide fragment is highlighted. Shown are representative peptide fragments with increased or decreased deuterium exchange with ligand (*) or without changes in deuterium exchange with ligand (–). The peptic peptides cover the majority of the sequence of reptin with the numbering shown below the amino acid sequence. The sequence includes six residual N-terminal amino acids from the "tag" after precision protease cleavage from glutathione beads, amino acids and includes GPLGST (Fig. S7a and b†). (d) The reptin protein (PDB 3UK6) is displayed as a cartoon in green. The ADP molecule from reptin is shown as spheres in cyan. Regions with suppressed deuterium exchange are shown in red and regions with increased deuterium exchange are shown in blue. Key regions which form a dimer interface (R428-S439, arrow 1) and the ATP pocket (T81-G95 (arrow 2) & Y362-C375 (arrow 3)) are highlighted. The most dominant peptide fragments which show alterations in deuterium exchange ((c) and Table S2†) map either around the ADP binding site or at the dimer interface.

This journal is © The Royal Society of Chemistry 2015

Open Access Article. Published on 20 March 2015. Downloaded on 29/06/2017 13:01:39.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

docking program LIDAEUS[25] was used to dock a conformer virtual library of 4.4 million compounds. The results were ranked based on the LIDAEUS score and the top 49 971 compounds were redocked using Vina and Autodock (Fig. S1†). A "rank-by-rank" consensus protocol prioritized hits, culminating in the selection of 30 compounds for assessment using an ELISA assay[17] with the peptide 104-FVLLNLVY-111 from the known reptin binding protein AGR2[17] (Table S1†). Hits from this assay were defined as compounds that modified (increased or decreased) significantly the signal corresponding to reptin binding to the AGR2-peptide compared to control. Of the 30 compounds tested (Fig. S2†), compound **1** (Fig. 1a) led to the most dramatic response and was therefore prioritized for study. Compound **1** contains a biphenyl substituent that is predicted to sit deep in the Walker A pocket where the adenine of ADP/ATP binds and a pyridine–oxazolo ring system which is predicted to extend out of the pocket (Fig. 1b and S3†).

Structure activity relationship and hydrogen–deuterium exchange studies

The synthesis of **1** (Scheme S1†) provided sufficient material for hit validation studies. An SAR study was then carried out to improve the activity of **1**. Modified analogs were either purchased or synthesized (Schemes S1 and S2 and Fig. S4†). The bromo-analog of **1**, now called Liddean (Fig. 1a), was found to be the most active analog. The results from this SAR study (see Fig. S5† legend for a more detailed discussion), supported by hydrogen–deuterium exchange (HDX) studies (Fig. 1c and d),[26] were consistent with the proposed binding mode. Importantly, suppression of HDX by Liddean was most pronounced for amino acids 362–375 in the Walker A site (Fig. S6†). Interestingly, an increased rate of HDX was observed along the length of the α-helix (81–95) that contacts to the Walker A site. In addition, deuterium exchange was suppressed for amino acids 428–439 which are located at the known protein–protein interface in reptin homodimers (Fig. 1c and d). It was therefore decided to assess whether Liddean had an effect on the oligomerization status of reptin.

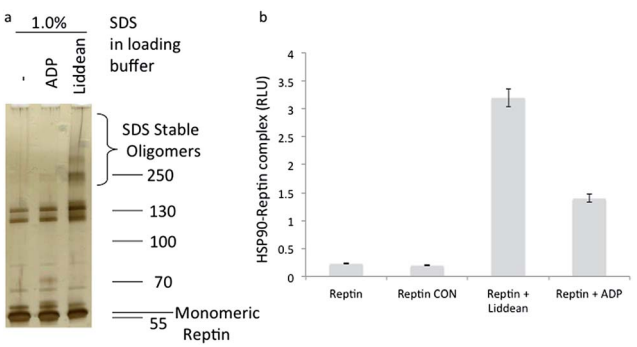


Fig. 2 (a) The effect of Liddean on the oligomerization dynamics of reptin. (a) Reptin (1 μg) was subjected to denaturing SDS (0.1%) gel electrophoresis in ADP binding buffer without or with ADP (100 μM) or Liddean (100 μM), as indicated. After 30 minutes of incubation at room temperature, gel loading buffer was added (with 1% SDS concentration), and electrophoresis was then carried out. Reptin protein was visualized by silver staining. See Fig. S8† for additional data. (b) The effect of Liddean on the interaction of reptin with the molecular chaperone HSP90. SBP-tagged HSP90 was captured on streptavidin coated wells and reptin (100 ng) was added in the ligand free and ligand bound states (100 μM where indicated). After washing the amount of reptin bound was quantified using an anti-reptin polyclonal antibody and peroxidase conjugated-anti rabbit IgG. The data are plotted as relative reptin activity (in RLU) as a function of the ligand bound state of reptin. CON = control.

Liddean alters reptin's oligomerization status

Oligomerization of reptin is known to increase in the presence of bound ligands ATP or ADP.[24] Reptin was subjected to SDS denaturing (0.1%) electrophoresis in the presence of varying amounts of SDS (Fig. 2a and S8a†). The preincubation of reptin with Liddean induced the formation of a stable oligomeric form of reptin (MW ≈ 250 kDa, Fig. 2a) as well as additional bands, corresponding to higher order oligomers (possibly hexamers based on the apparent mass). These data suggested that

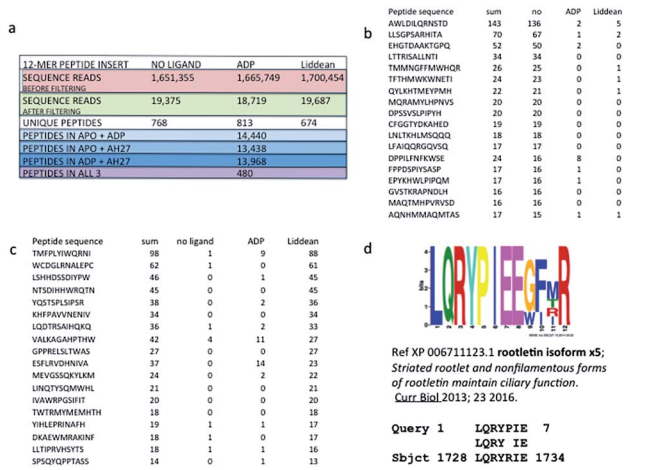


Fig. 3 Discovery of new Liddean-dependent interaction motifs for reptin. Next generation sequencing of peptide-phage pool obtained from a reptin screen in the *apo* and ligand bound state. Reptin was captured onto the solid phase without or with ligands ADP or Liddean. After selection of the peptide library on reptin protein, elution and propagation in bacteria, the phage DNA was amplified using PCR primer sets that capture the sequences flanking the peptide insert (as in Fig. S9†). Pooling of all phage into deep sequencing reactions can be done with subsequent deconvolution using the "bar code" whose position in the primer is indicated. (a) Parameters from the sequencing reactions from a representative screen are summarized. These include: (i) the sequencing reads before filtering non-specific binding peptides; (ii) the number of sequencing reads in *apo* or ligand bound protein; (iii) and the number of peptides that are shared in a number of *apo* or ligand bound screens. (b and c) Representative peptides that are enriched in the ligand bound state or suppressed in the ligand bound state are indicated to highlight a representative set of raw sequencing reads. (d) An example ciliopathy protein present in the list of human proteins which contain consensus sites identified by our Liddean-bound reptin screen. Processing the top 500 peptides from the *apo* and Liddean bound reptin using MEME to identify the top 10 consensus motifs (http://meme.nbcr.net/meme/cgi-bin/meme.cgi) highlights the distinct sets of motifs acquired in the *apo* and ligand bound form. The motifs were processed using MAST or *blastp* to identify targets in the human proteome that have matches to these motifs, some of which are listed as potential ciliopathy targets.

Open Access Article. Published on 20 March 2015. Downloaded on 29/06/2017 13:01:39.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Liddean was able to modify reptin's oligomerization dynamics with reptin being present in a more stable oligomeric state in the presence of Liddean. Liddean was also more effective at stabilizing reptin than ADP (Fig. 2a) with as little as 2 μM Liddean inducing stable oligomers (Fig. S8b,† lane 2). An attempt to correlate Liddean's ability to induce reptin homo-oligomerization and its stimulation of reptin's binding to the AGR2 peptide proved successful (Fig. S8†).

The link between nucleotide-induced changes in oligomerization status and function is a key feature of AAA+ proteins including reptin. To assess whether our Liddean-induced change in oligomerization status led to a modification of reptin's PPIs, we initially evaluated the effect of Liddean on reptin's known PPI with the molecular chaperone HSP90. Streptavidin-mimetic tagged (SBP) HSP90 was captured on solid phase followed by the addition of either apo-, ADP- or Liddean-bound reptin. Liddean (and to a lesser extent ADP) increased the stability of the reptin–HSP90 complex (Fig. 2b). Encouraged by the fact that clear changes could be observed in the presence of Liddean, we next used it as a tool to examine Liddean-induced changes on the global peptide-binding space of reptin.

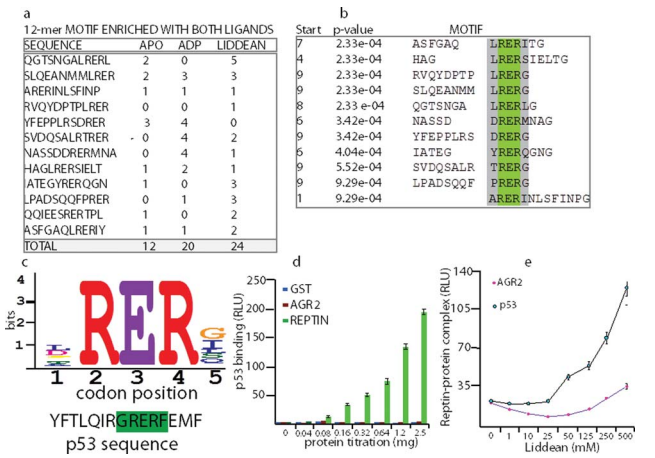


Fig. 4 Identification and validation of p53 as a novel reptin interacting protein. (a) 1000 peptides that were enriched in the ligand bound state (including those shown) were processed using MEME to identify consensus motifs (http://meme.nbcr.net/meme/cgi-bin/meme.cgi). (b) The panel represents data from a 12-*mer* peptide screen where the core motif identified is highlighted as ϕRERϕ or LRER[L/G]. (c) A blast motif screen using the MEME derived peptide consensus sites gave rise to a peptide derived from the tumor suppressor protein p53. (d) GST-tagged reptin and AGR2 proteins were assayed for their ability to bind to full length p53 using ELISA. p53 protein was absorbed onto the solid phase, and the indicated proteins were titrated in solution phase. The binding was detected using GST antibodies, followed by peroxidase conjugated secondary antibodies and processing using chemiluminescence. The data is plotted as binding activity as a function of protein amount (in RLU). (e) The effect of Liddean on the stability of the reptin–AGR2 and reptin–p53 protein interactions was evaluated. Either p53 or AGR2 were absorbed onto the solid phase and reptin (200 ng) was added in 50 μl of buffer containing increasing amounts of the indicated ligand. After 60 minutes incubation at room temperature, reptin protein bound to its target was quantified as indicated in the methods (in RLU).

Discovering new Liddean-stimulated peptide docking motifs on reptin using next generation sequencing of a phage-peptide combinatorial library

A combinatorial peptide-based selection assay exploiting next-generation "deep" DNA sequencing of peptide-phage pools was carried out using reptin in its ligand-free, ADP- or Liddean-bound form. Peptides were processed through 3 rounds of biopanning using a 12-*mer* peptide library, and peptide pools were sequenced (Fig. 3a, S9 and Table S3†). Comparison of the peptides identified using apo-reptin with those obtained when Liddean-bound reptin was used showed that, as expected, the binding of some peptides to reptin were suppressed (for representative raw peptide reads see Fig. 3b) or elevated by Liddean (Fig. 3c). Recent data has shown that reptin has an important interaction with the cytosolic cilia machinery.[27,28] This is a new cellular interactome for reptin distinct from its known links to the chaperone and transcription systems. It is interesting to note that several ciliopathy proteins are present in

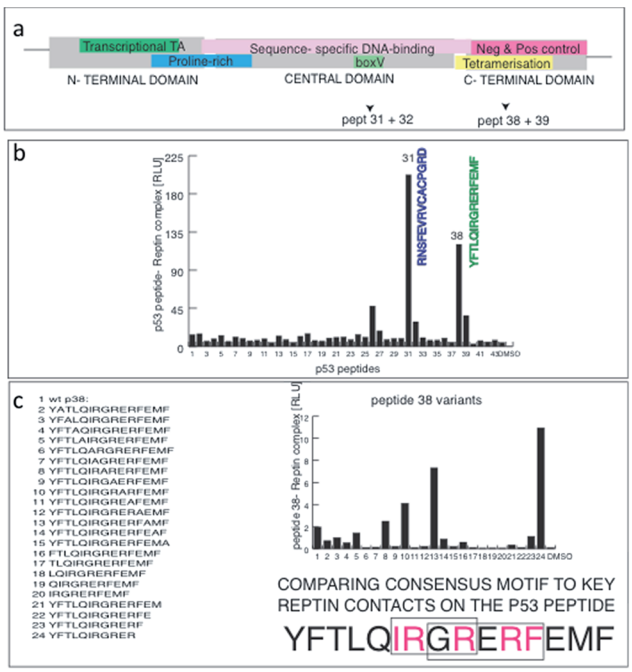


Fig. 5 Fine mapping of the dominant linear peptide docking site of reptin on p53. (a) The domain structure of p53 including the sites bound by MDM2 (in green), proline rich motif (blue), the specific DNA-binding domain (pink); tetramerization domain (yellow); and the C-terminal regulatory domain (in red). The arrows highlight the two binding sites mapped for reptin (in panels below). (b) An overlapping series of synthetic biotinylated peptides derived from the open reading frame of p53 were captured on streptavidin coated solid phase and reptin binding was measured as indicated in the methods. Two domain regions from p53 bound to reptin and mapped to the BOX-V domain (RNS…GRD) and to the tetramerization domain (YFT…EMF). The latter peptide contain two repeats of the ϕRERϕ or LRER[L/G] motifs identified from the ligand responsive peptides using MEME (Fig. 4). (c) Alanine scan mutagenesis of peptide 38 identifies important amino acids for reptin binding to p53; in the core sequence IRGRERFEMF, mutating IRGR or the overlapping RERF motif abrogates reptin binding to p53. This functional alanine mutagenesis is consistent with the MEME derived peptide motif from the deep sequencing.

This journal is © The Royal Society of Chemistry 2015

Open Access Article. Published on 20 March 2015. Downloaded on 29/06/2017 13:01:39.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

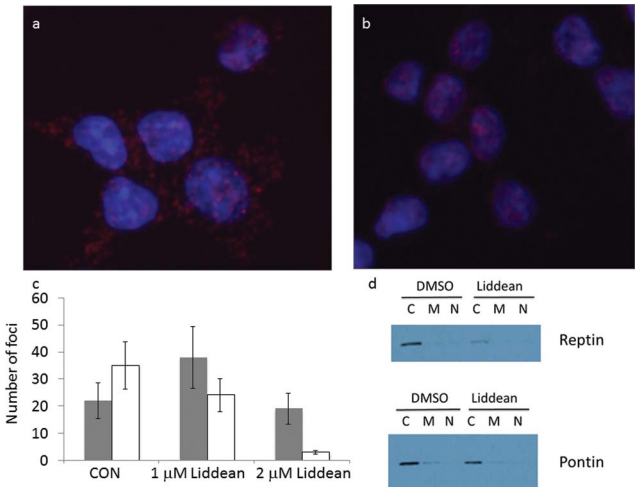


Fig. 6 The effects of Liddean on reptin–pontin interactions in cell models using the proximity ligation assay. HCT116 cells were processed using the proximity ligation method to identify whether reptin forms a PPI in cells and the images are superimposed using DAPI to highlight the nuclear (blue) or cytosolic foci location. The data highlight the foci of: (a) reptin–pontin in DMSO treated cells; (b) reptin–pontin foci in cells treated with 2 μM Liddean. (c) Number of cytoplasm (black bar) and nuclear (white bar) foci in the absence of presence of Liddean (1 μM and 2 μM). See Fig. S11† for the image of HCT116 cells in the presence of 1 μM Liddean; (d) immunoblots that show the amount of reptin or pontin after chemical fractionation into cytoplasmic, mitochondrial, and nuclear fractions.

the list of human proteins which contain consensus sites identified by our Liddean-bound reptin screen (Fig. 3d for one example, and Fig. S10†). Whilst the identified binding sites have yet to be validated, these data are entirely consistent with this approach being able to deliver a molecular peptide-binding "fingerprint" for reptin. In addition, the change in binding motifs identified in the presence of Liddean, linked with this ligand's observed effect on the oligomerization of reptin, provides indirect evidence for a substantial allosteric effect on the peptide-binding profile of reptin.

During these studies, several consensus peptide motifs were identified that were enriched in both ADP and Liddean bound forms. Our attention was drawn to the enrichment of motifs containing a ϕRERϕ sequence by ADP and/or Liddean (Fig. 4a and b). When this motif was compared with the human proteome, a motif was identified in the tetramerization domain of the p53 tumor suppressor (Fig. 4c). An ELISA assay demonstrated that the reptin protein was able to bind to purified human p53 whereas GST and AGR2 did not (Fig. 4d). Liddean was shown to stabilize the p53–reptin complex, consistent with the results from peptide-phage display (Fig. 4e).

In order to fine map the reptin binding site on p53, overlapping biotinylated peptides from human p53 (Fig. 5a and b) were probed with reptin to evaluate binding. Two dominant peptides (Fig. 5b), one which overlapped with the known MDM2 binding site in the central domain of p53 (peptide 31) and one with a motif in the tetramerization domain of p53 (peptide 38) bound to reptin. This provides at least two docking sites for reptin on p53. Alanine scanning mutagenesis of peptide 38 revealed that the key amino acid contacts required for efficient reptin binding to p53 form the core ϕRERϕ motif (Fig. 5c). These data further confirmed that our peptide combinatorial screen linked with next-generation sequencing can identify dominant docking sites for reptin on substrates and that Liddean can induce changes in reptin binding activity to important proteins such as AGR2, HSP90, ciliary proteins and p53.

Evaluation of the effects of Liddean on reptin–pontin interactions in a cell-based assay

In the final part of this study we assessed whether Liddean could be used to study reptin function in cells. To do this reptin's interaction with its most dominant paralog, pontin was observed using proximity ligation assays[29] in HCT116 cells. The complex between reptin and pontin was found to be largely cytoplasmic in the absence of Liddean (Fig. 6a; DMSO only). By contrast, Liddean (used at up to 2 μM) led to a substantial reduction in the cytoplasmic reptin–pontin foci and a change from clear punctate nuclear foci to aggregated reptin–pontin foci (Fig. 6b, c and S11†). Similar results were observed in p53-null cells (Fig. S12a–c†). Reptin and pontin expression levels were found to be equivalent with or without Liddean with the majority of both proteins being in the cytoplasmic fractions, relative to the mitochondrial and nuclear fractions using chemical fractionation (Fig. 6d). However, since the nucleus can be leaky in such chemical fractionation experiments, we also evaluated reptin and pontin proteins using immunofluorescence (IF; Fig. S12d–g†). The total amounts of reptin or pontin proteins also remained relatively unchanged in the absence or presence of Liddean (Fig. S12d–g†) and as defined by immunoblotting using urea lysis buffer (Fig. S12h and i†). Although the IF demonstrates a largely cytosolic reptin pool (Fig. S12d and f†), pontin does indeed show mixed cytosolic and nuclear pools (Fig. S12e and g†) suggesting the chemical fractionation does induce a leaching of pontin into the cytosol. These data together indicate that Liddean does not induce global changes in reptin or pontin protein levels or localization, but instead stimulates the production of reptin–pontin complexes in the nucleus. Together with the previous experiments, these data validate Liddean as a novel chemical tool with which to probe both *in vitro* and *in vivo* changes in the functions of the AAA+ protein reptin.

Discussion

The AAA+ proteins, including reptin, are known for their nucleotide binding sites, oligomeric propensity, and wide range of functions.[5] Here we built on the observation that nucleotide binding is known to regulate the oligomerization status of reptin. In particular, we were interested in exploring whether by modifying reptin's oligomeric state we would change the proteins it interacts with.[30] A number of assays could have been used to screen for novel reptin ligands, including helicase activity, ATPase activity, fragment based drug discovery, and/or high throughput competitive binding with nucleotide ligands.[31] Indeed, a recent approach has identified small molecule

Open Access Article. Published on 20 March 2015. Downloaded on 29/06/2017 13:01:39.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

inhibitors of the ATPase activity of the reptin paralog, pontin.[32] Instead we used an *in silico* screen to predict small molecules that would target the deep nucleotide binding pocket (Walker A site) that is a characteristic feature of oligomeric AAA+ proteins. The identified hit (compound **1**) was subjected to SAR analysis guided by an ELISA based PPI assay to generate the chemical tool, Liddean. The SAR data and results from hydrogen–deuterium (H–D) exchange experiments supported our proposal that Liddean binds in the Walker A site of reptin. In addition, Liddean stabilised higher order oligomers of reptin as evidenced by (i) the estimated mass using a denaturing gel electrophoretic screen and (ii) the use of HDX experiments that revealed suppression of deuterium incorporation at the previously reported dimer interface.

Reptin is known to interact with a variety of chromatin and chaperonin signalling proteins and is often considered important in oncogenesis.[33] More recently, based on genetic screens, reptin has been linked to the assembly of cilia structures on the plasma membrane.[27,28] With Liddean in hand, we decided to assess whether it could be used to find and provide details about reptin PPIs. As a proof of principle we confirmed that Liddean-bound reptin modified reptin's known interaction with HSP90. Next we decided to use the stabilized, oligomeric state of Liddean-bound reptin *in vitro* to search for "consensus peptide motifs" using a combination of next generation deep-sequencing and a combinatorial peptide-phage library. Whilst a very information rich dataset was obtained, our focus turned to a novel reptin-interacting motif in the p53 tumor suppressor protein. Independent screens verified that indeed reptin can bind to p53-derived peptides mainly through a peptide motif that is located in the p53 tetramerization domain (Fig. 4 and 5). As a further demonstration of the utility of our approach, it was also noted that peptide motifs in proteins of the cytosolic cilia machinery were identified. On-going studies will probe the details of these proposed interactions.

Whilst Liddean had proved a very useful tool *in vitro*, we wanted to assess whether it could also be used in cells. A proximity ligation assays[29] was therefore used to assess the effect of Liddean on the interaction of reptin with its dominant partner, pontin, in cells. This technique enabled us to observe directly the reptin–pontin complex in the cytoplasm of cancer cells. Interestingly, reptin formed significantly more nuclear foci with pontin on addition of Liddean. Whilst an explanation for this observed redistribution remains challenging, dramatic changes in protein expression levels or localization have been ruled out.

Conclusions

In discovering and subsequently using Liddean, our chemical biology platform has provided novel insights into the PPIs associated with the important human protein reptin. Ultimately, identifying reptin's complete PPI network and explaining how the network is controlled is central to understanding its role in normal and disease processes. An important concept relating to PPIs is that small linear peptide motifs can form dynamic and specific docking sites for a protein.[34–36] Small molecule stabilization/destabilization of these motifs provides a promising approach towards overall modulation of protein function.[1] Indeed, there are PPI drug leads emerging that are being applied in the clinic; the most notable of which targets the linear peptide motif-binding groove of the MDM2 oncoprotein.[37] As dynamic linear peptide-motif based PPIs form a vast untapped landscape in biology and medicine,[38] approaches that facilitate the discovery of such interactions will provide new avenues to impact on drug discovery programs.[39] We believe that the approach we have outlined in this report is applicable to other members of the AAA+ superfamily.

Materials and methods

Protein–protein interaction assays of reptin

The expression and purification of reptin protein was carried out as previously described[17] and as discussed in Fig. S13† with the following exceptions. The cells were initially put into a buffer containing 50 mM HEPES pH 8.0 and 10% sucrose before being snap frozen, the rest of the components of the lysis buffer were then added with the exception of Triton X-100 which was added at a 0.1% concentration rather than 0.5%. The lysate was incubated with glutathione beads for 150 minutes at 4 °C with rotation before the washes were carried out. The protease used to cleave the reptin from the beads was HRV 3C. Biotinylated AGR2 derived peptides (or p53 where indicated) were coated overnight onto streptavidin coated wells and reptin binding was measured in buffers as described previously.[17] All biotinylated peptides were obtained from Chiron Mimotopes (Australia). When small molecules were evaluated to the indicated final concentrations (balanced with DMSO carrier), reptin was added immediately to the reaction well to allow binding competition to take place in the presence of the AGR2 (or p53) peptide. The wells were washed[17] and bound reptin was detected using a reptin antibody coupled to anti-rabbit secondary antibody and chemiluminescence. Binding activity was quantified by chemiluminescence using a Fluoroskan Ascent FL Labsystems. For measuring the effects of ligands on the SDS-resistant oligomerization state of reptin using SDS gel electrophoresis, reptin protein (1 μg) was added to buffer B (25 mM HEPES, pH 8.0; 10% glycerol, 10 mM KCl, 1 mM DTT) with the indicated amounts of ADP or Liddean. Following incubation at room temperature for 60 minutes, samples were processed for electrophoresis as indicated in the Fig. S8† legend.

Hydrogen–deuterium exchange mass spectrometry

Deuteration of the reptin either free or in complex with Liddean was initiated by a sequential dilution into deuterated water with 0.1% DMSO final concentration. The molar ratio between reptin and Liddean was 1 : 5 (as summarised in Fig. S7†). The exchange was done at 21 °C and was quenched by the addition of 1 M HCl in 1 M glycine at 1 min and 5 minutes followed by rapid freezing in liquid nitrogen. Each sample was quickly thawed and injected onto an immobilized pepsin column (15 μl bed volume, flow rate 20 μl min^{-1}, 0.1% formic acid in water). Peptides were trapped and desalted on-line on a peptide

 This journal is © The Royal Society of Chemistry 2015

Open Access Article. Published on 20 March 2015. Downloaded on 29/06/2017 13:01:39. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

microtrap (Michrom Bioresources, Auburn, CA) for 1 minute at a flow rate 20 μl min^{-1}. The peptides were eluted onto an analytical column (Jupiter C18, 1.0 × 50 mm, 5 μm, 300 Å, Phenomenex, CA) and separated by a linear gradient. The injection, switching valves, immobilized pepsin column, trap cartridge, and the analytical column was kept at 1 °C in a cool box (within the robotics system). Mass spectrometric analysis was done on an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) with ESI ionization on line connected with a robotic system based on a HTS-XT platform (CTC Analytics company). The instrument was operated in a data-dependent mode for peptide mapping (HPLC-MS/MS). Each MS scan was followed by MS/MS scans of the top three most intensive ions from both CID and HCD fragmentation spectra. Tandem mass spectra were searched using SequestHT against the cRap protein database (ftp://ftp.thegpm.org/fasta/cRAP) containing sequence of reptin protein. Sequence coverage was visualized with Proteome Discoverer 1.4 software (Thermo Fisher Scientific). Analysis of deuterated samples was done in HPLC-MS mode with ion detection in the orbital ion trap and data were processed in HDX Workbench. Graphs showing deuteration kinetics were plotted by DrawHDXPlot (MSTools).

Combinatorial peptide phage screen

Peptide phage was carried out using the 12-*mer* Ph.D.™ Phage display library (New England Biolabs). The surface panning procedure (direct target coating) was carried out as instructed by the manufacturer's protocol with an additional protein capture step. The micro titer wells were coated as directed with a rabbit anti-reptin polyclonal antibody overnight. The wells were then washed three times with tris buffered saline with 0.1% Tween-20 (TBST). The wells were then blocked with 3% bovine serum albumin in TBS for 1 hour. Washes were once again carried out before the addition of reptin with and without ligand (either ADP or Liddean at 100 μM) and incubated for one hour at RT. Washes were then carried out 6 times with TBST, containing ligand in the ligand treated wells. The phage pool was then added, again containing ligand if required, and incubated for 1 hour. Non-binding phage was removed and the wells were washed 10 times with PBST prior to the elution of the phage as directed by the manufacturer. Amplification and titering of the phage was also carried out with each round of panning as per the manufacturer's instructions. Titering was carried out to ensure that the phage pool being panned was not greater than that of the original pool once it had been amplified and to check for white-type lytic phage contamination. Polymerase chain reaction (PCR) and deep sequencing of phage was carried out in the following stages; (i) PCR was used to amplify phage DNA from each round of screening using the primer bar codes in table code (Table S3†) that have an Illumina adaptor sequence and a 3 letter bar code; (ii) equal amounts of DNA was gel purified on a 2% agarose gel to create a pool (5 μg) that was sequenced by Otogenetics (USA). Fastq files were then captured using a custom tool programmed in Java language (script available upon request) that was used to extract amino acid sequences from raw NGS reads. Only forward reads were processed (as reverse reads do not capture the bar code). Barcode and mimotope DNA sequences were extracted from reads that passed quality control based on exact match search for bordering sequences. All sequences having nonsense (not in list) bar code were filtered out. Mimotope sequences having inappropriate length or containing nonsense codons (stop codons as well as some other "forbidden" codons that should not be present according to New England BioLabs phage library manual) were filtered out. Sequences passing these filters were translated, grouped by resulting peptide sequence and sorted as indicated in the tables.

Duo-link proximity ligation assay

Cells were grown on coverslips until they reached around 30% confluency upon which the small molecule Liddean was transfected into cell and incubated for 24 hours (transfection was performed using 1 or 2 μM molecule with DMSO control balanced in 100 μl of DMEM containing attractene carrier). Cells were fixed with 4% paraformaldehyde dissolved in PBS and permeabilized using 0.25% Triton X-100 in PBS. Duolink II (green) assay from Olink Bioscience was carried out following manufactures instructions. Briefly, unspecific antibody binding was blocked by the addition of 3% BSA (w/v) in PBS for 30 minutes at room temperature. Primary antibodies were diluted in 3% BSA (w/v) in PBS and added to the cells overnight at 4 °C. Negative controls include no primary or no secondary antibodies which gave rise to no foci in subsequent assay processing (data not shown). PLA probes were added to cells and incubated to 1 hour at 37 °C. Ligation mix was added for 30 minute at 37 °C followed by the polymerase amplification mix for 2 hours at 37 °C. Coverslips were mounted in S3023 Mounting medium (Dako). Results were visualized using a BX51 (Olympus) fluorescent microscope. Antibodies and PLA probes include; anti-reptin/TIP49B/RUVB2 rabbit antibody ab36569 (abcam) (1 : 250 dilution); monoclonal anti-pontin 5G3-11 (Sigma) (1 : 250 dilution); Duolink II PLA probe anti rabbit PLUS (Olink) (1 : 10 dilution); and Duolink II PLA probe anti mouse MINUS (Olink) (1 : 10 dilution).

Acknowledgements

Financial support for this project was provided by Cancer Research UK (CRUK grant C21383/A6950). CRUK C483/A10706 and C483/A8033; EPSRC EP/F500421/1 doctoral training centre in cell and proteomic technologies; GACR P206/12/G151 and the state budget of the Czech Republic (LO1413).

Notes and references

1 H. Jubb, A. P. Higueruelo, A. Winter and T. L. Blundell, *Trends Pharmacol. Sci.*, 2012, **33**, 241–248.
2 Y.-C. Chen, S. V. Rajagopala, T. Stellberger and P. Uetz, *Nat. Methods*, 2010, **7**, 667–668.
3 J.-P. Lambert, G. Ivosev, A. L. Couzens, B. Larsen, M. Taipale, Z.-Y. Lin, Q. Zhong, S. Lindquist, M. Vidal, R. Aebersold,

Open Access Article. Published on 20 March 2015. Downloaded on 29/06/2017 13:01:39.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

T. Pawson, R. Bonner, S. Tate and A.-C. Gingras, *Nat. Methods*, 2013, **10**, 1239–1245.
4 P. Tompa, *Nat. Chem. Biol.*, 2012, **8**, 597–600.
5 J. Snider, G. Thibault and W. A. Houry, *Genome Biol.*, 2008, **9**, 216.
6 B. M. Stinson, A. R. Nager, S. E. Glynn, K. R. Schmitz, T. A. Baker and R. T. Sauer, *Cell*, 2013, **153**, 628–639.
7 N. Nano and W. Houry, *Philos. Trans. R. Soc., B*, 2013, DOI: 10.1098/rstb.2011.0399.
8 X. Shen, G. Mizuguchi, A. Hamiche and C. Wu, *Nature*, 2000, **406**, 541–544.
9 Z. O. Jonsson, S. K. Dhar, G. J. Narlikar, R. Auty, N. Wagle, D. Pellman, R. E. Pratt, R. Kingston and A. Dutta, *J. Biol. Chem.*, 2001, **276**, 16279–16288.
10 T. Ikura, V. V. Ogryzko, M. Grigoriev, R. Groisman, J. Wang, M. Horikoshi, R. Scully, J. Qin and Y. Nakatani, *Cell*, 2000, **102**, 463–473.
11 Y. Doyon, W. Selleck, W. S. Lane, S. Tan and J. Cote, *Mol. Cell. Biol.*, 2004, **24**, 1884–1896.
12 N. J. Krogan, M. C. Keogh, N. Datta, C. Sawa, O. W. Ryan, H. Ding, R. A. Haw, J. Pootoolal, A. Tong, V. Canadien, D. P. Richards, X. Wu, A. Emili, T. R. Hughes, S. Buratowski and J. F. Greenblatt, *Mol. Cell*, 2003, **12**, 1565–1576.
13 Y. Cai, J. Jin, L. Florens, S. K. Swanson, T. Kusch, B. Li, J. L. Workman, M. P. Washburn, R. C. Conaway and J. W. Conaway, *J. Biol. Chem.*, 2005, **280**, 13665–13670.
14 M. A. Wood, S. B. McMahon and M. D. Cole, *Mol. Cell*, 2000, **5**, 321–330.
15 A. Bauer, S. Chauvet, O. Huber, F. Usseglio, U. Rothbacher, D. Aragnol, R. Kemler and J. Pradel, *EMBO J.*, 2000, **19**, 6121–6130.
16 J. H. Kim, B. Kim, L. Cai, H. J. Choi, K. A. Ohgi, C. Tran, C. Chen, C. H. Chung, O. Huber, D. W. Rose, C. L. Sawyers, M. G. Rosenfeld and S. H. Baek, *Nature*, 2005, **434**, 921–926.
17 M. M. Maslon, R. Hrstka, B. Vojtesek and T. R. Hupp, *J. Mol. Biol.*, 2010, **404**, 418–438.
18 N. Izumi, A. Yamashita, A. Iwamatsu, R. Kurata, H. Nakamura, B. Saari, H. Hirano, P. Anderson and S. Ohno, *Sci. Signaling*, 2010, **3**, ra27.
19 A. S. Venteicher, Z. Meng, P. J. Mason, T. D. Veenstra and S. E. Artandi, *Cell*, 2008, **132**, 945–957.
20 S. Gorynia, T. M. Bandeiras, F. G. Pinho, C. E. McVey, C. Vonrhein, A. Round, D. I. Svergun, P. Donner, P. M. Matias and M. A. Carrondo, *J. Struct. Biol.*, 2011, **176**, 279–291.
21 P. M. Matias, S. Gorynia, P. Donner and M. A. Carrondo, *J. Biol. Chem.*, 2006, **281**, 38918–38929.
22 M. Petukhov, A. Dagkessamanskaja, M. Bommer, T. Barrett, I. Tsaneva, A. Yakimov, R. Quéval, A. Shvetsov, M. Khodorkovskiy, E. Käs and M. Grigoriev, *Structure*, 2012, **20**, 1321–1331.
23 K. L. Y. Cheung, J. Huen, W. A. Houry and J. Ortega, *Biochem. Cell Biol.*, 2010, **88**, 77–88.
24 A. Niewiarowski, A. S. Bradley, J. Gor, A. R. McKay, S. J. Perkins and I. R. Tsaneva, *Biochem. J.*, 2010, **429**, 113–125.
25 P. Taylor, E. Blackburn, Y. G. Sheng, S. Harding, K.-Y. Hsin, D. Kan, S. Shave and M. D. Walkinshaw, *Br. J. Pharmacol.*, 2008, **153**, S55–S67.
26 L. Hernychova, P. Man, C. Verma, J. Nicholson, C.-A. Sharma, E. Ruckova, J. Y. Teo, K. Ball, B. Vojtesek and T. R. Hupp, *Proteomics*, 2013, **13**, 2512–2525.
27 V. Stolc, M. P. Samanta, W. Tongprasit and W. F. Marshall, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 3703–3707.
28 L. Zhao, S. Yuan, Y. Cao, S. Kallakuri, Y. Li, N. Kishimoto, L. DiBella and Z. Sun, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 12697–12702.
29 C.-M. Clausson, A. Allalou, I. Weibrecht, S. Mahmoudi, M. Farnebo, U. Landegren, C. Wählby and O. Söderberg, *Nat. Methods*, 2011, **8**, 892–893.
30 R. Queval, C. Papin, M. Dalvai, K. Bystricky and O. Humbert, *J. Biol. Chem.*, 2014, **289**(49), 33999–34012.
31 J. Rosenbaum, S. Baek and A. Dutta, *Sci. Signaling*, 2013, **6**, 1–6.
32 J. Elkaim, M. Lamblin, M. Laguerre, J. Rosenbaum, P. Lestienne, L. Eloy, T. Cresteil, F.-X. Felpin and J. Dessolin, *Bioorg. Med. Chem. Lett.*, 2014, **24**, 2512–2516.
33 A. Grigoletto, P. Lestienne and J. Rosenbaum, *Biochim. Biophys. Acta*, 2011, **1815**, 147–157.
34 K. Van Roey, B. Uyar, R. J. Weatheritt, H. Dinkel, M. Seiler, A. Budd, T. J. Gibson and N. E. Davey, *Chem. Rev.*, 2014, **114**, 6733–6778.
35 P. Tompa, N. E. Davey, T. J. Gibson and M. M. Babu, *Mol. Cell*, 2014, **55**, 161–169.
36 M. L. Miller, L. J. Jensen, F. Diella, C. Jorgensen, M. Tinti, L. Li, M. Hsiung, S. A. Parker, J. Bordeaux, T. Sicheritz-Ponten, M. Olhovsky, A. Pasculescu, J. Alexander, S. Knapp, N. Blom, P. Bork, S. Li, G. Cesareni, T. Pawson, B. E. Turk, M. B. Yaffe, S. Brunak and R. Linding, *Sci. Signaling*, 2008, **1**, ra2, DOI: 10.1126/scisignal.1159433.
37 I. Ray-Coquard, J.-Y. Blay, A. Italiano, A. Le Cesne, N. Penel, J. Zhi, F. Heil, R. Rueger, B. Graves, M. Ding, D. Geho, S. A. Middleton, L. T. Vassilev, G. L. Nichols and B. N. Bui, *Lancet Oncol.*, 2012, **13**, 1133–1140.
38 M. J. Basse, S. Betzi, R. Bourgeas, S. Bouzidi, B. Chetrit, V. Hamon, X. Morelli and P. Roche, *Nucleic Acids Res.*, 2013, **41**, D824–D827.
39 H. Dinkel, K. Van Roey, S. Michael, N. E. Davey, R. J. Weatheritt, D. Born, T. Speck, D. Krüger, G. Grebnev, M. Kuban, M. Strumillo, B. Uyar, A. Budd, B. Altenberg, M. Seiler, L. B. Chemes, J. Glavina, I. E. Sánchez, F. Diella and T. J. Gibson, *Nucleic Acids Res.*, 2014, **42**, D259–D266.

This journal is © The Royal Society of Chemistry 2015

Journal of Proteomics 199 (2019) 89–101

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

An inter-subunit protein-peptide interface that stabilizes the specific activity and oligomerization of the AAA+ chaperone Reptin

Dominika Coufalova[a,1], Lucy Remnant[b,1], Lenka Hernychova[a], Petr Muller[a], Alan Healy[c], Srinivasaraghavan Kannan[d], Nicholas Westwood[c], Chandra S. Verma[d,e,f], Borek Vojtesek[a,2], Ted R. Hupp[a,b,g,*,2], Douglas R. Houston[h,2]

[a] Regional Centre for Applied Molecular Oncology, Masaryk Memorial Cancer Institute, 656 53 Brno, Czech Republic
[b] University of Edinburgh, Institute of Genetics and Molecular Medicine, Edinburgh, Scotland EH4 2XR, United Kingdom
[c] St Andrews University, St Andrews, Scotland, United Kingdom
[d] Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, Matrix 07-01 138671, Singapore
[e] School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, 637551, Singapore
[f] Department of Biological Sciences, National University of Singapore, 14, Science Drive 4, 117543, Singapore
[g] University of Gdansk, International Centre for Cancer Vaccine Science, ul. Wita Stwosza 63, 80-308 Gdansk, Poland
[h] University of Edinburgh, Institute of Quantitative Biology, Biochemistry and Biotechnology, Edinburgh, Scotland EH9 3BF, United Kingdom

A B S T R A C T

Reptin is a member of the AAA+ superfamily whose members can exist in equilibrium between monomeric *apo* forms and ligand bound hexamers. Inter-subunit protein-protein interfaces that stabilize Reptin in its oligomeric state are not well-defined. A self-peptide binding assay identified a protein-peptide interface mapping to an inter-subunit "rim" of the hexamer bridged by Tyrosine-340. A Y340A mutation reduced ADP-dependent oligomer formation using a gel filtration assay, suggesting that Y340 forms a dominant oligomer stabilizing side chain. The monomeric Reptin^{Y340A} mutant protein exhibited increased activity to its partner protein AGR2 in an ELISA assay, further suggesting that hexamer formation can preclude certain protein interactions. Hydrogen-deuterium exchange mass spectrometry (HDX-MS) demonstrated that the Y340A mutation attenuated deuterium suppression of Reptin in this motif in the presence of ligand. By contrast, the tyrosine motif of Reptin interacts with a shallower pocket in the hetero-oligomeric structure containing Pontin and HDX-MS revealed no obvious role of the Y340 side chain in stabilizing the Reptin-Pontin oligomer. Molecular dynamic simulations (MDS) rationalized how the Y340A mutation impacts upon a normally stabilizing inter-subunit amino acid contact. MDS also revealed how the D299N mutation can, by contrast, remove oligomer de-stabilizing contacts. These data suggest that the Reptin interactome can be regulated by a ligand dependent equilibrium between monomeric and hexameric forms through a hydrophobic inter-subunit protein-protein interaction motif bridged by Tyrosine-340.

Significance: Discovering dynamic protein-protein interactions is a fundamental aim of research in the life sciences. An emerging view of protein-protein interactions in higher eukaryotes is that they are driven by small linear polypeptide sequences; the linear motif. We report on the use of linear-peptide motif screens to discover a relatively high affinity peptide-protein interaction for the AAA+ and pro-oncogenic protein Reptin. This peptide interaction site was shown to form a 'hot-spot' protein-protein interaction site, and validated to be important for ligand-induced oligomerization of the Reptin protein. These biochemical data provide a foundation to understand how single point mutations in Reptin can impact on its oligomerization and protein-protein interaction landscape.

1. Introduction

The AAA+ superfamily contains a group of ATP-regulated hexameric proteins that are highly conserved in evolution [1]. Their dynamic oligomeric properties provide a model system to understand allosteric regulation of polypeptide assembly and function [2]. The RuvBl1/2 orthologues, also named Reptin and Pontin, represent two representative AAA+ protein members. Structural studies have defined the properties of complexes of homo-oligomeric Pontin [3], hetero-dodecameric Pontin and Reptin [4] and homo-hexameric Reptin [5]. There are two ATP-binding motifs in such AAA+ proteins: the Walker A and Walker B ATP-binding motifs that mediate ATP binding and hydrolysis respectively.

Reptin's key evolutionarily conserved activity, from yeast to vertebrates, regulates chromatin-modifying machinery, including Ino80 [6,7], and SCRAP [8]. Additional protein-protein interactions from studies in human cancer include those with ATF2 [9], Myc [10], transcription associated protein β-catenin [11], tumor suppressor ARF [12],

* Corresponding author at: Regional Centre for Applied Molecular Oncology, Masaryk Memorial Cancer Institute, 656 53 Brno, Czech Republic.
E-mail address: ted.hupp@ed.ac.uk (T.R. Hupp).
[1] Joint first authors.
[2] Joint corresponding authors.

https://doi.org/10.1016/j.jprot.2019.02.012
Received 24 October 2018; Received in revised form 16 January 2019; Accepted 23 February 2019
Available online 09 March 2019
1874-3919/ © 2019 Published by Elsevier B.V.

and the oncogenic protein AGR2 [13]. Reptin and Pontin are also regulators of all PIKK (phosphatidylinositol 3-kinase-related protein kinase) members [14] that in turn mediate the DNA damage response and nonsense-mediated mRNA decay (NMD). Reptin and Pontin also regulate the assembly of the telomerase holoenzyme through interactions with TERT and dyskerin [15]. A major paradigm in Reptin structure and function was identified using yeast models where it forms a multi-subunit chaperonin complex with HSP90 [16].

Reptin is a potential drug target in cancer due to its oncogenic nature and large ATP-binding pocket that might make it suitable for drug-discovery programmes [17]. As such, we had previously performed in silico screening of a large chemical library to identify several lead molecules that can interact with the Walker A ATP-binding site [18]. The screen resulted in the small molecule *Liddean* which: (i) mimicked ATP or ADP in the stimulation of the oligomerization of Reptin protein; (ii) regulated Reptin's heterologous protein-protein interactions; and (iii) stimulated Reptin assembly with Pontin in cell-based systems. Together these observations suggested that occupation of the ATP-pocket of Reptin by a small molecule ATP-mimetic like *Liddean* can stimulate its oligomerization. In this report, we focus on understanding the role of ligand binding in promoting key protein-protein contacts at the oligomerization interface of Reptin during its self-oligomerization. We identified an inter-subunit hotspot in the 'rim' of the hexamer at Tyrosine-340 that plays an important role in ligand-induced oligomerization that promotes hexamer stability and regulates Reptin's specific activity.

2. Experimental procedures

2.1. Site directed mutagenesis of Reptin at D299N and at Y340A

All chemicals were obtained from Sigma unless otherwise indicated. Human Reptin was cloned into Gateway Entry clone (Invitrogen) for subsequent use. The human Reptin sequence for cloning into a new *E. coli* expression system was amplified using the following primers: forward primer 5′-GGGACAAGTTTGTACAA AAAAGCAGGCTTCCTGGAA GTTCTGTTCCAGGGGCCC ATGGCAACCGTTACAGCCACAACC-3′ and reverse primer 5′-GGGGACCACTTTGTACAAGAAAGCTGGGTCCAGGA GGTGTCCATGGTCTCG-3′. The forward primer had a *PreScission* protease cleavage site inserted. Following amplification, the polymerase chain reaction (PCR) product was first inserted into pDONR201 and then into pDEST-15 using Gateway technology (*Invitrogen*) to generate GST-tagged Reptin. Point mutations in the above mentioned plasmids were introduced using the following primers: for the Reptin D299N mutant, forward primer 5′-GAGTGCTGTTCATCAACGAGGTCCACA TGC-3′ and reverse primer 5′-GCATGTGGACCTCGTTGATGAACAGCA CTC-3′; and for the Reptin Y340A mutant, forward 5′- GAATCCGGGG CACCAGCGCCCAGAGCCCTCACGGCA-3′ and reverse 5′- GCCGTGAGG GCTCTGGGCGCTGGTGCCCCGGATTC-3″.

2.2. Expression and purification of proteins

2.2.1. Reptin WT and mutant production

Reptin proteins were expressed as indicated previously [13]. Briefly, plasmids encoding the indicated Reptin proteins were transformed into BL21-AI (Thermo Fischer Scientific). The cells were grown in LB medium (1 L) at 37 °C up to an A_{600} of 0.5. Induction of gene expression was achieved by adding arabinose to the culture (final concentration 0.2%). The bacterial culture was grown at room temperature (21 °C) for another 3 h and then pelleted by centrifugation. The lower temperature of induction (rather than 30 °C or 37 °C) gave rise to reproducible recover of Reptin protein. Cells were lysed with 30 ml of a buffer containing 10% sucrose, 50 mM HEPES, pH 8.0, 400 mM NaCl, 0.1% Triton X-100, 1 mM DTT, 1 mM benzamidine, 0.5 mg/ml lysozyme, and protease inhibitors for 30 min on ice and then sonicated for 3 × 5 s bursts to reduce the viscosity. The clarified bacterial lysate was obtained by centrifugation for 15 min at 15,000 g. The lysate was incubated with glutathione beads (0.5 ml) (GE Healthcare) for 150 min at 4 °C with rotation in a 50 ml conical tube. The beads were thoroughly washed three times by mixing the beads with 10 bead-volumes of a high salt buffer (20 mM HEPES, pH 7.5, 1 M NaCl, 1 mM DTT, and 1 mM benzamidine) to remove non-specifically bound proteins. The beads were then washed three times with 10 bead-volumes of a low salt buffer (20 mM HEPES, pH 7.5, 0.05 M NaCl, 1 mM DTT, and 1 mM benzamidine). The beads were finally washed three times with 10 bead-volumes of PreScission buffer (GE Healthcare, Uppsala, Sweden) and then resuspended in 1 bead volume (0.5 ml) of PreScission buffer. GST-tagged PreScission protease (GE Healthcare; 5 units) was added to the slurry and incubated overnight rotating in the cold room. The beads were sedimented at 2340 g for one minute and the supernatant containing cleaved Reptin protein was recovered. The purity of Reptin was confirmed using SDS-gel electrophoresis as described previously [13]. The beads were washed with an additional column volume of PreScission buffer to collect additional, but more diluted protein. The first fraction from the elution, which is more concentrated than the second fraction, was invariably used. The Reptin protein cleaved off the column using PreScission protease (GE Healthcare) was stored at 4 °C at concentrations from 0.5 mg/ml to 1 mg/ml. The protein was stable in this condition for at least 6 months.

2.2.2. AGR2 purification

His-tagged AGR2 (from [19]) was expressed in BL21-AI and purified using Ni^{2+}-nitrilotriacetic acid-agarose (Qiagen). according to the manufacturer's instructions. In detail, the cells were grown in LB medium (0.5 L) at 37 °C up to an A_{600} of 1.0 and protein was induced by adding arabinose to the culture (0.2%). The culture was grown at 30 °C for 3 h and then pelleted by centrifugation. The pellets were lysed with 40 ml of a buffer containing 20 mM Tris–HCl, pH 8.0, 150 mM NaCl, 10 mM $MgCl_2$, 0.05% Tween 20, 10% glycerol, 20 mM imidazole, pH 8.0, and 0.5 mg/ml lysozyme for 30 min on ice. The lysate was then sonicated for 3 × 5 s bursts to reduce the viscosity. The clarified bacterial lysate was obtained by centrifugation for 15 min at 15,000 g. The lysate was incubated with Ni-agarose on rotary shaker at 4 °C for 1 h to capture the AGR2 protein. The beads were washed two times in lysis buffer and three times in lysis buffer with 40 mM imidazole, and then the AGR2 protein was eluted with lysis buffer containing 150 mM imidazole. Purity was confirmed using SDS-gel electrophoresis as described [19]. AGR2 protein eluted from the column was stored at −70 °C at concentrations from 1 to 3 mg/ml. The protein was stable in this condition for at least 2 years.

2.2.3. HSP90 purification

The coding sequence of the human HSP90AA1 gene (Hsp90α, NM_001017963.2) was cloned into a pDEST17 vector containing an N-terminal His 6 × -tag cleavable by TEV protease. All cloned genes were expressed in BL21(DE3) RIPL cells. The cells were grown in LB medium at 37 °C up to an A_{600} of 0.5. Induction of gene expression was achieved by adding isopropyl β-D-thiogalactopyranoside to the culture (final concentration 1 mM). The bacterial culture was grown at 30 °C for another 3–4 h and then pelleted by centrifugation for 15 min at 6000 *g*. The cells were resuspended in His binding buffer II (50 mM Tris, pH 8.0, 0.2 M NaCl, 2 mM $MgCl_2$, 10% glycerol). Cell suspensions were enriched with lysozyme (1 mg/ml) and PMSF (1 mM) and then sonicated. Bacterial lysates were obtained by centrifugation for 30 min at 12,000 g. His 6 -tagged Hsp90α was purified using a HisTrap column, and the His 6 x-tag was cleaved by overnight incubation with TEV protease. The protein was subsequently exchanged into His binding buffer II and then subjected to a second immobilized metal affinity chromatograph to remove His 6 x-tag and His 6-TEV. The flow-through fractions were concentrated and further processed by gel filtration using a HiPrep 16/60 Sephacryl S-200 HR column (GE Healthcare). The purity of all isolated proteins was confirmed by SDS-PAGE/Coomassie

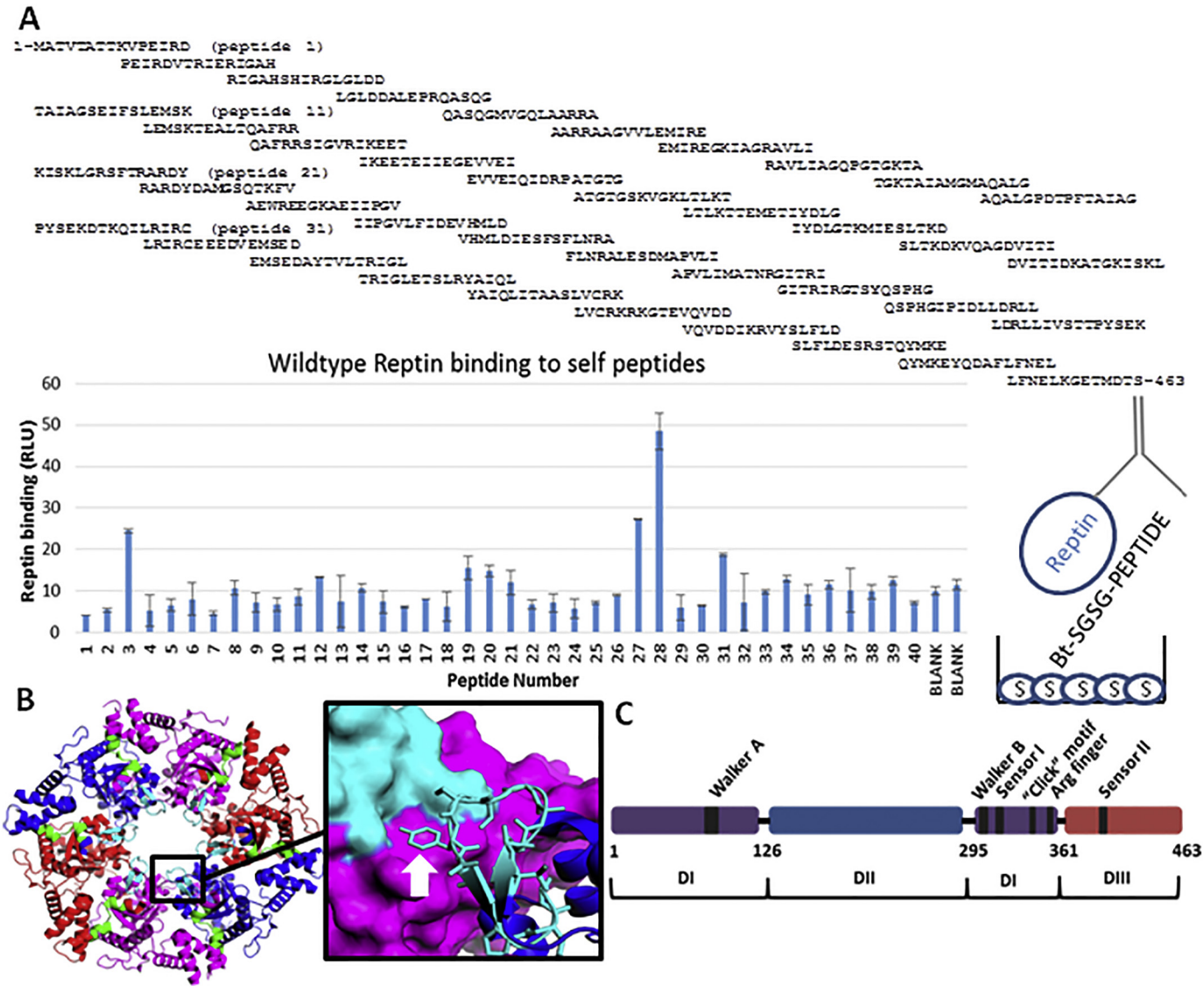


Fig. 1. Structure of Reptin (RUVBL2). (A) ELISA assay measuring the binding of wt-Reptin to self-peptides. The bottom right diagram highlights the order of addition in the Reptin binding reaction, including; (i) coating the ELISA wells with Streptavidin (S); (ii) adding the biotinylated peptide containing a SGSG space; (iii) the addition of Reptin protein; and (iv) the addition of the primary anti-reptin antibody (Y shape; in grey) that is detected using HRP conjugated secondary antibody. The top panel summarized the overlapping, biotinylated peptides derived from the Reptin open reading frame (from the N-terminus (peptide 1) through the C-terminus (peptide 40) were coated onto streptavidin ELISA wells. After washing, Reptin was added and binding measured using anti-Reptin antibody in conjunction with HRP labeled secondary antibody. Reactions were monitored via chemiluminescence. Binding data is represented by RLU (relative light units) as a function of peptide number. (B) X-ray crystal structure of the Reptin hexamer (PDB id: 3UK6) with the subunits shown in different colors (red, magenta, blue). An inter-subunit interface (428–439) is colored green and the oligomerization site (333–345) is colored aqua. (C) Details of the oligomerization site (aqua) between two Reptin monomers (blue and magenta). The side chain of the Tyrosine finger (Y340) is located in the oligomerization pocket of the second Reptin. (D). Reptin structure. A linear rod structure of Reptin domain organization. Domain DI (in purple) containing highly conserved motifs of AAA+ proteins such as the Walker A, the Walker B, the Sensor I, the Arg finger and the "Click" (inter-subunit stabilizing) motif described in this article. Domain DIII (in red) contains the conserved motif of AAA+ protein region named Sensor II. The insertion domain DIII is also highlighted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

staining [20]. All proteins were finally exchanged into final assay buffers using 7-kDa molecular mass cutoff Zeba spin desalting column (Thermo Fischer).

2.2.4. Reptin ELISA protein-peptide interaction assays

Overlapping biotinylated peptides were obtained from Chiron Mimotopes (Australia). Peptides derived from Reptin are displayed in the Fig. 1 legend. Reptin protein binding to biotinylated protein was measured using ELISA. Fifteen-mer peptides with a 5-mer overlap (Fig. 1A) and containing a N-terminal biotin followed by a SGSG spacer were purchased from Chiron Mimotopes. Peptides were dissolved at 5 mg/ml in DMSO. In peptide binding assays, ELISA plates were coated with streptavidin (AnaSpec) overnight at 37 °C and washed with PBS-Tween-20 (0.1% Tween-20). The biotinylated peptides diluted in water (1 μg per well) were added for 1 h followed by blocking with BSA 3% in PBS 1×. Reptin diluted in blocking buffer was then added to the plate and incubated for a further 1 h. When ligands such as ADP or *Liddean* were used, the chemicals were added to buffer containing Reptin for 30 min prior to addition of the reactions to the ELISA wells to allow sufficient oligomerization of the protein induced by the ligands. Blocking buffer and DMSO only controls were included. After further washes, wells were incubated with the Reptin primary antibody and

then horseradish peroxidase conjugated secondary antibody for 1 h each at room temperature followed by further washing and ECL addition and the luminescence produced was measured. The data are presented as Relative Light Units and represent triplicates graphed as a mean with error bars representing the deviations from the mean.

2.2.5. Gel filtration

Reptin and the indicated mutants (Y340A and D299N; 25 μg) were incubated without ligand or with ligand (1 mM ADP) in 100 μl of HEPES buffer (pH 7.6) containing PBS. Samples were injected onto a 25 ml Superose-12 gel filtration column in the same buffer and fractions were collected at a flow rate of 0.5 ml/min. In the ligand bound state, the column was pre-incubated in the same buffer but including 1 mM ADP. Gel filtration markers were used to calibrate the column and to demonstrate ADP did not alter the elution of marker proteins (data not shown). Samples eluting from the column (50 μl) of each fraction were absorbed onto ELISA wells overnight in a cold room and wells were incubated with PBS with BSA (3%) to block reactive interaction sites. The wells were washed with PBS (containing 0.1% Tween-20) (0.1%) followed by incubation with the anti-Reptin antibody (1:1000; Abcam) in PBS-tween-20 (0.1%) including BSA (3%). Reactions were processed with HRP-conjugated secondary antibodies (DAKO; 1:1000) to define the elution profile of Reptin (in RLU, chemiluminescence). The elution profiles are highlighted in Fig. 3. Molecular mass markers (GE Healthcare) that calibrated the column were injected (100 μl) at 10 mg/ml concentrations (in deionized water) and elution peak volumes were defined by absorbance ($_{max}$280 nm).

2.3. Experimental design and statistical rationale for hydrogen-deuterium exchange mass spectrometry

2.3.1. Sample preparation of Reptin proteins

Reptin, either wt or mutant (5 μM), was incubated with 1 mM ATP for 1 h at 21 °C prior to the exchange. The exchange was initiated by a sequential dilution into deuterated buffer (50 mM Tris-HCl pD 7.1, 150 mM NaCl, 1 mM EDTA, 2 mM $MgCl_2$ and 1 mM DTT) to a final concentration of 2 μM. The sequential dilution took approximately 25 s. Counting of the incubation time was initiated after addition of the first aliquot of deuterated buffer. For sample with ATP the deuterated buffer was enriched with 1 mM ATP. The exchange was carried out at room temperature and was quenched by the addition of 1 M HCl in 1 M glycine at 5 min followed by rapid freezing in liquid nitrogen.

2.3.2. Digestion and HPLC separation

Each sample was thawed and injected onto an immobilized pepsin column (15 μl bed volume, flow rate 20 μl/min, 2% acetonitrile/0.05% trifluoroacetic acid). Peptides were trapped and desalted on-line on a peptide microtrap (Michrom Bioresources, Auburn, CA) for 2 min at flow rate 20 μl/min. Next, the peptides were eluted onto an analytical column (Jupiter C18, 1.0 × 50 mm, 5 μm, 300 Å, Phenomenex, CA) and separated using a 2 min linear gradient elution of 10%–40% B buffer in A buffer, followed by 31 min isocratic elution at 40% B. Solvents used were: A buffer – 0.1% formic acid in water, B buffer – 80% acetonitrile/0.08% formic acid. The immobilized pepsin column, trap cartridge and the analytical column were kept at 1 °C.

2.3.3. Mass spectrometry and data analysis

Mass spectrometric analysis was carried out using an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) with ESI ionization on-line connected with a robotic system based on the HTS-XT platform (CTC Analytics, Zwingen, Switzerland). The instrument was operated in a data-dependent mode for peptide mapping (HPLC-MS/MS). Each MS scan was followed by MS/MS scans of the top three most intensive ions from both CID and HCD fragmentation spectra. Tandem mass spectra were searched using SequestHT against the cRap protein database (ftp://ftp.thegpm.org/fasta/cRAP) containing the sequence of the Reptin protein with the following search settings: mass tolerance for precursor ions of 10 ppm, mass tolerance for fragment ions of 0.6 Da, no enzyme specificity, two maximum missed cleavage sites and no-fixed or variable modifications were applied. The false discovery rate at peptide identification level was set to 1%. Sequence coverage was analyzed with Proteome Discoverer software version 1.4 (Thermo Fisher Scientific) and graphically visualized with MS Tools application [21]. Totally 94% of Reptin sequence was covered by 249 identified peptides (Supplementary Fig. 1). Analysis of deuterated samples was done in HPLC-MS mode with ion detection in the orbital ion trap. The MS raw files together with the list of peptides (peptide pool) identified with high confidence characterized by requested parameters (retention time, XCorr, and charge) were processed using HDExaminer version 2.2 (Sierra Analytics, Modesto, CA). The software analyzed protein and peptides behavior, created the uptake plots that showed peptide deuteration over time with calculated confidence level (high, medium confidence are accepted, low confidence is rejected). The HDX data in Figs. 5 and 7 are representative of data obtained in at least three independent experiments. The results from peptide pools (Supplementary Tables 1-3) were displayed as butterfly comparison plots (Supplementary Figs. 2-4) that showed the deuteration percentage as a function of peptide index calculated for different protein states. This function includes other useful information about how well the peptide pool covers protein (number of peptides, percentage of coverage protein sequence, standard deviation, average redundancy etc., Supplementary Figs. S2-S4). Other graphs showing the evolution of deuteration at individual parts of the protein at the same time and different protein states were plotted using the GraphPad Prism version 5.03 for Windows (GraphPad Software, San Diego, CA, USA). The bimodal isotopic envelopes were analyzed by HX-Express 2 [22]. Molecular structures were rendered using PyMOL [23]. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE repository with the dataset identifier “PXD008226”; Username: reviewer05996@ebi.ac.uk, Password: I7pjGxG1.

2.4. Modeling and molecular dynamics (MD) simulations

MD simulations of *apo* and ADP bound monomeric and hexameric Reptin in its wt, Y340A and D299N mutant forms were carried out with the *pemed.CUDA* module of the program Amber14 [24]. The structures of the mutants (Y340A, D299N) were generated by substituting Tyr340 with Ala and Asp299 with Asn in the wild-type protein [5]. All atom versions of the Amber 14SB force field (ff14SB) [25] and the generalized Amber force field (GAFF) [26] were used to represent the protein and ADP respectively. The *Xleap* module was used to prepare the system for the MD simulations. Each structure was solvated in an octahedral box with TIP3P [27] water molecules, leaving at least 10 Å between the solute atoms and the borders of the box. All the simulation systems were neutralized with appropriate numbers of counterions. MD simulations were carried out in explicit solvent at 300 K. During the simulations, the long-range electrostatic interactions were treated with the particle mesh Ewald [28] method using a real space cutoff distance of 9 Å. The Settle [29] algorithm was used to constrain bond vibrations involving hydrogen atoms, which allowed a time step of 2 fseconds during the simulations. Solvent molecules and counterions were initially relaxed using energy minimization with restraints on the protein and inhibitor atoms. This was followed by unrestrained energy minimization to remove any steric clashes. Subsequently the system was gradually heated from 0 to 300 K using MD simulations with positional restraints (force constant: 50 kcal mol^{-1} $Å^{-2}$) on protein and inhibitors over a period of 0.25 nseconds allowing water molecules and ions to move freely. During an additional 0.25 nseconds, the positional restraints were gradually reduced followed by a 2 nseconds unrestrained MD simulation to equilibrate all the atoms. For each system, three independent MD simulations (assigning different initial velocities) were carried out for 250 nseconds with conformations saved every 10

pseconds. Figures were drawn using PyMOL [23].

3. Results

3.1. Defining hotspot Reptin interaction sites using an overlapping biotinylated Reptin self-peptide library array

Reptin forms multiple inter-subunit protein-protein interaction contacts. However, the dominating, functional interaction sites or "hotspots" that stabilize inter-subunit protein-protein interactions [30] involved in Reptin homo-oligomer assembly are not well-defined. The emerging view of protein interactions in higher eukaryotes highlights the concept that the large portion of protein-protein interactions are driven by linear peptide motifs that encompass hotspots for stabilizing domain-domain interactions [31]. We used this concept as an approach to determine whether Reptin exhibits self-binding activity towards linear peptide motifs derived from the Reptin amino acid sequence. We set up an overlapping biotinylated peptide scan assay using sequences derived from Reptin (Fig. 1A) to identify self-peptides that might interact stably with the protein. Identification of such linear peptides could point out high affinity, functional, "hotspots" in inter-subunit protein-protein docking site.

Synthetic biotinylated peptides contained a SGSG amino acid spacer followed by the appropriate overlapping Reptin amino acid sequence are listed in Fig. 1A. The biotinylated peptides were captured on streptavidin coated ELISA wells and unbound peptides was washed away. Monomeric Reptin was added to measure binding activity. There was only one major, dominating self-peptide binding activity that was mapped to peptide 28 (Fig. 1A; Biotin-SGSG-GITRIRGTSYQSPHG) which resides at an inter-subunit region of the Reptin oligomer (Fig. 1B). This peptide forms a tyrosine finger which embeds into each adjacent subunit in the hexameric structure (Fig. 1C). We name this tyrosine-containing peptide a click motif. Interestingly, deuterium exchange mass spectrometry was used previously to demonstrate that the Reptin ligand and ADP-mimetic named *Liddean* suppressed deuteration at this site within the ring structure formed between all six subunits [18]. These data are consistent with a model that oligomerization induced by *Liddean* might suppress deuteration at this buried inter-subunit interface motif.

3.2. Ligands that induce Reptin oligomerization suppress its interaction with the click interface peptide 28

We next set up assays that measure the impact of a ligand that can induce Reptin oligomerization on its binding to peptide 28. A titration of Reptin protein (up to 320 ng) into ELISA wells coated with either peptide 27, peptide 28, or peptide 29 (Fig. 1A) confirmed that the protein exhibits a high preference for peptide 28 (Fig. 2A). This peptide contains the amino acid sequence with the bridging Tyrosine 340 amino acid side chain (Fig. 1C). Using this assay, we next determined whether Reptin binding to peptide 28 can be affected by ligands that induce its oligomerization. Using fixed amount of Reptin (500 ng), increasing concentrations of ADP from 1 μM to 200 μM resulted in a reduction in its binding to peptide 28 (Fig. 2B). Similarly, increasing the concentrations of *Liddean* from 1 μM to 200 μM resulted in suppression of Reptin binding to peptide 28 (Fig. 2B). As a control, the weak binding activity of Reptin observed to peptide 27 was also reduced by increasing the concentrations of either ADP or *Liddean* (Fig. 2B). These data are consistent with the hypothesis that Reptin homo-oligomer formation stabilized through the inter-subunit docking of the peptide 28 click motif to the adjacent subunit (as in Fig. 1B and C) would sterically preclude the self-peptide 28 binding *in trans*.

We next also evaluated previously identified Reptin interacting proteins, as controls, to determine how ligand-induced oligomerization impacts on Reptin binding activity. *Liddean* was used as an ATP mimetic ligand that binds to the Walker-A ATP binding domain [18]. ADP was also used to eliminate the contribution of ATPase activity of Reptin from the Walker B ATP-interaction site to its dynamic oligomerization. AGR2 is a previously published Reptin binding protein whose binding to Reptin has been fine-mapped [18]. As expected, we observed that both *Liddean* and ADP reduced Reptin binding to AGR2 (Fig. 2C).

Reptin is part of the R2TP complex and can interact with certain molecular chaperones, including HSP90 [32]. A titration of Reptin alone resulted in relatively low levels of binding to the HSP90 chaperone (Fig. 2C). The pre-incubation of Reptin with either *Liddean* or ADP (500 μM) increased the binding activity of Reptin (Fig. 2D). *Liddean* was more potent than ADP in stimulating the HSP90-Reptin complex (Fig. 2D). This is consistent with the increased potency of *Liddean* to induce SDS-stable oligomerization of Reptin [18]. These data suggest that the ligand bound form of Reptin (presumably oligomeric) or the ligand-free form of Reptin (presumably monomeric) have different affinities for partner proteins (summarized in Fig. 4).

3.3. Mutation of tyrosine-340 within the click peptide motif reduces ligand-stimulated oligomerization of Reptin

These data can be used to predict that if we mutate the Y340-peptide oligomerization interface (Fig. 1C) we might alter Reptin oligomerization. In addition, we might also impact on the affinity of Reptin for these proteins whose binding is regulated by Reptin oligomerization states (as in Fig. 2). In order to determine whether the Y340A mutation can impact on Reptin oligomerization states, we evaluated the relative mass of the wild-type and mutant Reptin by quantifying their elution profiles using gel filtration.

Two mutant forms of Reptin were purified and characterized compared to the wild-type protein. We chose to mutate Tyrosine 340 to Alanine ($Reptin^{Y340A}$), since the Tyrosine side-chain forms a possible key hydrophobic stabilizing interaction between monomers (Fig. 1C). We then asked how the loss of this apparently stabilizing tyrosine amino acid contact (Fig. 1C) can impact on ligand-dependent oligomerization. We also chose to mutate Aspartate 299 to Asparagine ($Reptin^{D299N}$). This mutation resides within the Walker B ATP binding domain and is implicated as a gain-of-function inhibitor of Reptin in cell growth since it has enhanced ATP-binding activity [13] because of a defective ATPase activity [33]. However, to our knowledge, despite significant functional cell based data on the biochemical characteristics of this D299N mutant, there has been no analysis of its oligomerization status in the absence and presence of ligand. As such, we included this mutant as another control to evaluate how the oligomerization of Reptin might be affected in comparison to the Y340A mutation.

An ELISA assay was used to measure the elution of Reptin from a Superose-12 gel filtration column and then to define its relative oligomerization state in the ligand-free and ligand bound forms. Following elution from the gel filtration column, aliquots of the fractions were absorbed onto 250 μl ELISA wells. Protein was detected using immunochemical detection as indicated in the Experimental Procedures. We would predict at the outset that wild-type Reptin would exhibit an increase in its oligomerization status in the presence of ligand. We would predict that the $Reptin^{Y340A}$ mutant would presumably be less oligomeric if our hypothesis is accurate that the Y340 side chain forms a stabilizing hotspot interaction with the adjacent subunit. We did not know at the outset how the $Reptin^{D299N}$ mutant would elute in the absence and presence of ligand.

We processed the samples without ADP or with ADP in the gel filtration buffer (1 mM), rather than ATP, as the use of ADP removes the contribution of ATP hydrolysis to the dynamic oligomerization state of the protein over the course of the gel filtration elution (90 min). Relative to molecular mass markers, wt-Reptin elutes as an apparent monomer on S12-sepharose (Fig. 3A). The preincubation of Reptin with ADP, and including ADP in the gel filtration buffer, results in an increase in a pool of the protein at a larger mass range suggesting the elution of a stable oligomer (a hexamer based on crystal structure data).

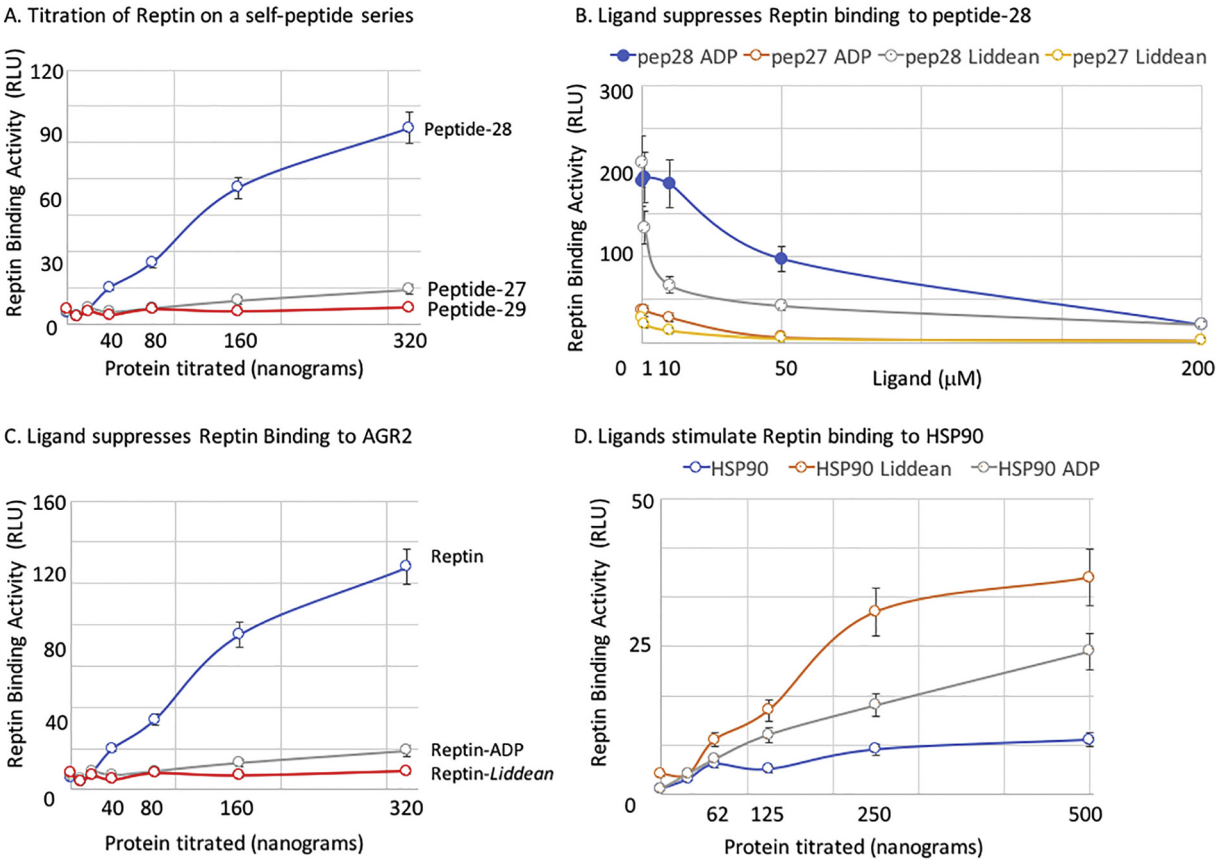


Fig. 2. Impact of ligands on the protein-binding function of Reptin. (A). ELISA wells pre-coated with streptavidin were incubated with the indicated peptides (27, 28, and 29) representing the motif surrounding the self-binding peptide (28; Fig. 1). Reptin was titrated to measure relative binding. Anti-Reptin and HRP-conjugated secondary antibodies were used to measure the extent of Reptin binding (in RLU, chemiluminescence). (B). ELISA wells pre-coated with streptavidin were incubated with the indicated peptides (27 and 28). Reptin (500 ng) was preincubated with or without increasing concentrations of the indicated ligand (DMSO), *Liddean*, or ADP) in buffer (50 μl). Anti-Reptin and HRP-conjugated secondary antibodies were used to measure the extent of Reptin binding (in RLU, chemiluminescence). (C). The previously published Reptin binding proteins (AGR2) was coated onto ELISA wells. Reptin was titrated with or without the indicated ligand (DMSO, ADP (100 μM), or *Liddean* (100 μM)). Anti-Reptin and HRP-conjugated secondary antibodies were used to measure the extent of Reptin binding (in RLU, chemiluminescence). (D). The previously identified Reptin interacting protein, HSP90, was coated onto ELISA wells. Reptin was titrated with or without the indicated ligand (DMSO, ADP (500 μM), or *Liddean* (500 μM)). Anti-Reptin and HRP-conjugated secondary antibodies were used to measure the extent of Reptin binding (in RLU, chemiluminescence).

We found that a reproducible proportion of the Reptin did not shift into the fully oligomeric (presumably hexameric) position, as highlighted by the presence of two peaks in the gel filtration elution profile (Fig. 3A). These data suggest there might be two distinct conformational pools of the recombinant Reptin purified from bacterial host (as suggested by bimodal hydrogen deuterium exchange data, see below). Alternatively, the gel filtration methodologies might dilute Reptin during the chromatography, reducing its concentration sufficiently below its association constant in the presence of ligand.

Elution of Reptin^{Y340A} protein resulted in a protein that mimicked wild-type Reptin profile in the absence of ligand but the inclusion of ADP in the buffer did not result in a shift it into a larger oligomeric state (Fig. 3B). These data suggest that the Y340 does provide a hotspot stabilization interaction between the subunits as its mutation precludes stable oligomer formation in the presence of the ligand. This is consistent with the ability of peptide-28 (containing the tyrosine 340 residue) to bind with a relatively high affinity to Reptin (Fig. 1A). As the unknown control, to our surprise, the ATP-binding gain-of-function mutant Reptin^{D299N} resulted in a profile (in the absence of ligand) that mimics wt-Reptin bound to ADP (Fig. 3C). These latter data suggest that the D299N mutant traps Reptin in an oligomeric state without ligand. These data are rationalized using molecular dynamic simulations (below) and we include this D299N mutant in our discussion on oligomer fluctuation as a comparison to the Y340A and wt forms of Reptin (Fig. 4). The data also suggest, although the D299N mutation has been defined as 'inactive' in cells due to its low ATPase activity, with respect to oligomerization, the D299N mutation induces a 'gain-of-function' property to the protein. However, since the protein is ATP hydrolysis inactive, it would presumably be not only gain-of-function with respect to oligomerization but dominant negative in its activity in cells.

Having established that the panel of wild-type and mutant proteins exhibit different oligomerization properties in the presence or absence of ligand, we next tested all in a range of biochemical assays to correlate oligomerization potential to its specific activity. This panel of Reptin proteins (wt-Reptin, Reptin^{Y340A}, Reptin^{S342A}, and Reptin^{D299N}) was next tested for binding to one of its most well-characterized binding proteins; the oncogenic AGR2 protein [13]. The binding of either wild-type Reptin or Reptin^{S342A} to AGR2 are relatively similar (Fig. 3D) suggesting that the Serine 342 site chain does not form a hot spot stabilizing interaction. Interestingly, Reptin^{D299N} exhibited a much lower specific activity in this AGR2 binding assay (Fig. 3D). Since Reptin^{D299N} exhibits enhanced oligomerization without ligand (Fig. 3C), these data are consistent with the prior data showing that that the ligand-bound form of Reptin has reduced AGR2 binding activity (Fig. 2D). By contrast, the Reptin^{Y340A} exhibited a much higher specific activity that wild-type Reptin in this AGR2 binding assay (Fig. 3D). These data are also consistent with the concept that Reptin can exist in dynamic oligomeric states. The monomeric state of the Y340A mutant (as suggested

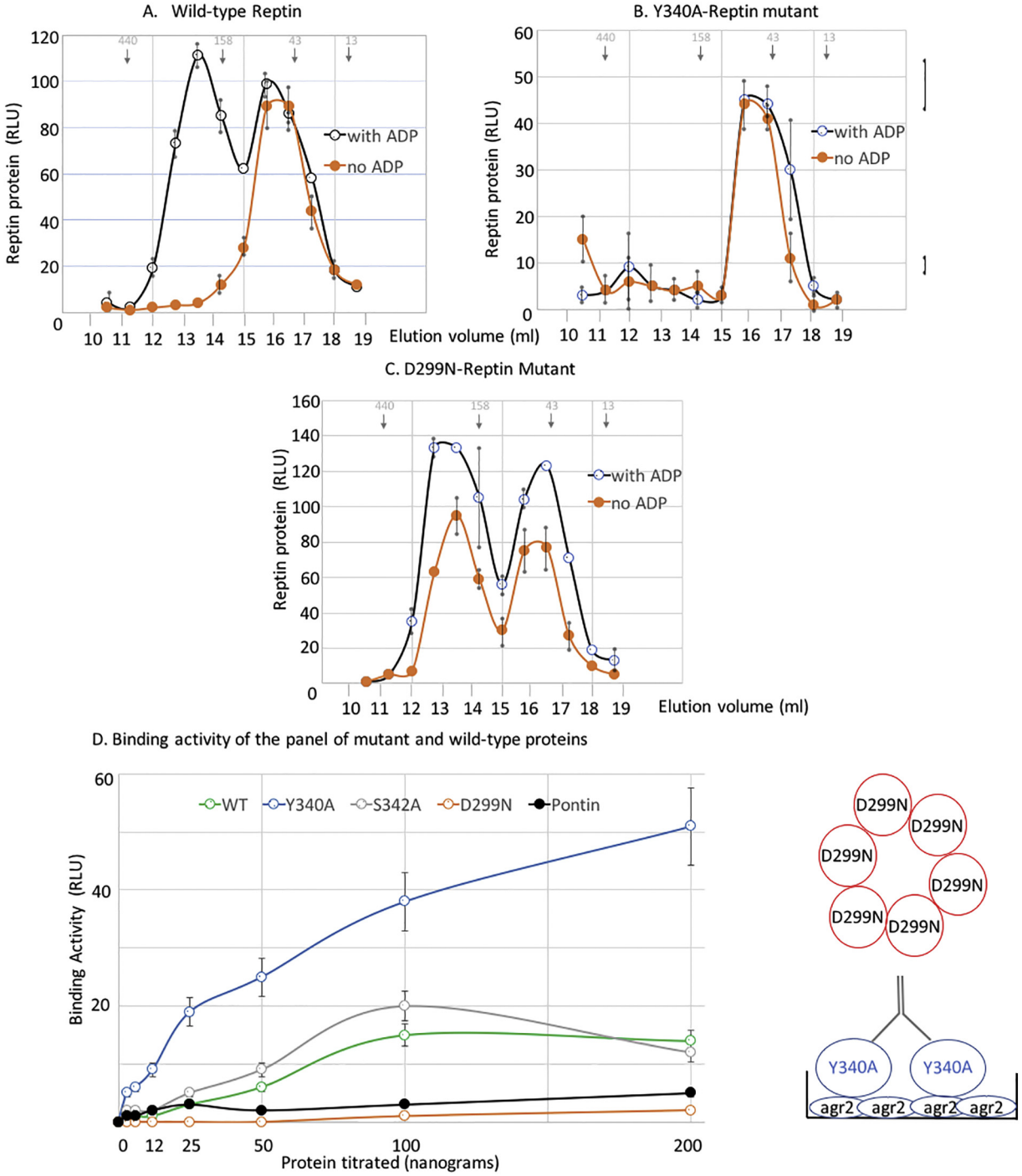


Fig. 3. The impact of mutation on Reptin quaternary structure and specific activity. (A). Wild-type Reptin, Reptin^{Y340A}, and Reptin^{D299N} were incubated with 1 mM ADP or without ligand and loaded onto a superose-12 gel filtration column (25 ml). Fractions of samples eluting from the column were absorbed onto ELISA wells. Anti-Reptin and HRP-conjugated secondary antibodies were used to measure the extent of Reptin elution and binding (in RLU, chemiluminescence). The arrows highlight the elution of molecular mass markers in kDa (440, Ferritin; 158, Aldolase; 43, Ovalbumin; 13, RNase A) in parallel processed in the absence of ADP. (B). Wild-type Reptin and the indicated mutants were analyzed in AGR2 protein binding assays. AGR2 (1 μg) was pre-coated onto ELISA wells. Reptin was titrated, and anti-Reptin and HRP-conjugated secondary antibodies were used to measure the extent of Reptin binding (in RLU, chemiluminescence). The diagram on the right interprets the data in the ELISA; mutations that drive Reptin into the monomeric state (Y340A) have a higher specific activity for AGR2, whilst mutations that stabilize Reptin in its oligomeric state (D299N) have a lower specific activity for AGR2 protein. These data are consistent with the attenuation of wild-type Reptin binding to AGR2 protein in the presence of oligomerization-promoting ligands, *Liddean* or ADP (Fig. 2).

by Fig. 3B) can increase its binding to some proteins such as AGR2 (Fig. 3D and Fig. 4).

3.4. Use of Hydrogen-deuterium exchange mass spectrometry (HDX-MS) to measure Reptin oligomerization dynamics

HDX-MS was used to measure the effect of ATP on the conformation of Reptin and the impact of the Y340A mutation on these properties.

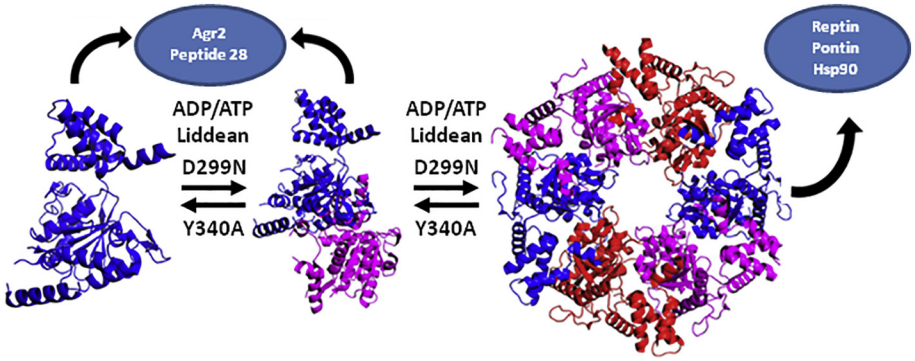


Fig. 4. A model describing the specific activity of Reptin as a function of its oligomeric state. Reptin is presented as existing in distinct dynamic oligomeric states. Mutations or ligands can impact on this flux. The reason for developing a staged model is, firstly; there is one stable inter-subunit protein-protein interface that can be detected as suppressed by HDX-MS in Reptin in the presence of ATP (Supplementary Fig. 5C and Fig. 6A). The wild-type Reptin and Y340A mutant form of Reptin exhibit similar suppression of deuteration at this inter-subunit protein-protein (Fig. 6A and B). These data suggest that the Y340A mutation has little impact on oligomer assembly at this site. By contrast, the Y340A mutation can attenuate full oligomerization at the 'rim' (Fig. 3B and Fig. 6B). By contrast to the Y340A mutation that attenuates oligomer formation, the D299N mutation stabilizes the hexameric state in the absence of ligand (Fig. 3C). The stabilization effect of the D299N mutation is rationalized in Figs. 7-9. In contrast to mutations, oligomerization (hexamerization) can be stimulated by ligands such as ATP, ADP, or *Liddean*. Oligomerization stimulation of Reptin can increase or decrease its binding to client proteins (Fig. 2). This is similar to a study on the assembly of Reptin and Pontin by Histone-3; where proteins that bind to the Reptin-Pontin system depend on the monomer to oligomer status of the two proteins [43].

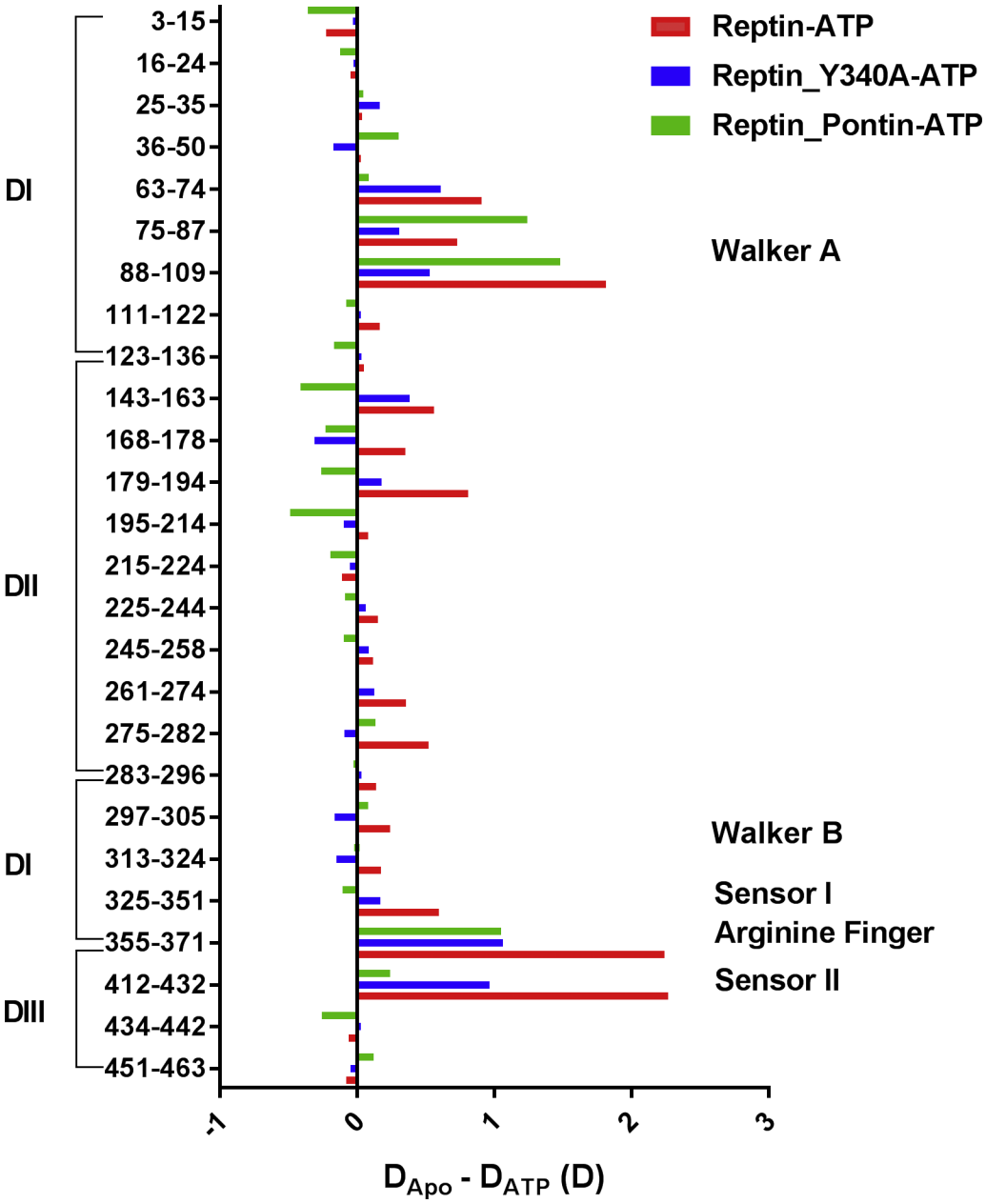


Fig. 5. Hydrogen-Deuterium exchange mass spectrometry of Reptin in differing conformational states. (A) wild-type Reptin; (B) Y340A mutant form of Reptin; (C) wild-type Reptin with Pontin. Reptin in all three conformational states (apo) was preincubated with or without ligand (ATP) and subjected to deuteration and pepsinization. The reaction products were subjected to mass spectrometry as indicated in the Experimental Procedures. The graphs highlight the change in the number of deuterons as a function of ligand from each isoform (See Supplementary Table 4). In addition, the average deuteration (%) of peptic peptides, as a function of primary amino acid sequence, of ligand-bound and ligand-free Reptin at the 300 s time point calculated as described [37] is summarized in Supplementary Fig. 6.

We have previously used HDX-MS methodology to define conformational changes induced upon mutation or ligand binding in proteins including MDM2 [34], CHIP [35], EpCAM [36], and HSP70 [37]. We first compared deuteration of wild-type Reptin with deuteration of wild-type Reptin pre-incubated with 1 mM ATP (Fig. 5 and Fig. 6A). Peptic peptide segments that exhibited different deuterium exchange rates at 300 s post-dilution with deuterated buffer as result of ATP binding are highlighted in Fig. 5 (in red). The data reflect changes in deuteron number as a function of the indicated peptic peptide fragments. These peptides are located in all three Reptin functional domains (Fig. 6A). In domain I (Fig. 1D and Fig. 6A), these segments contain amino acids 63–109 that comprise the ATP-binding Walker A motif (Supplementary Fig. 5; in white). This is consistent with ATP occupying the nucleotide binding pocket of Reptin under these conditions and serves as a positive control that the ATP-binding reaction was active under the deuteration dilution conditions. In addition, domain I containing amino acids 298–318, comprising the Walker B motif, also exhibited attenuated deuteration (Supplementary Fig. 5; in blue). This is also consistent with ligand interaction at this site leading to suppression of deuteration. A third motif in domain I includes amino acids 340–351 (Fig. 6A, in red) that contains the Tyrosine-340 residue and comprises the so-called “click” motif that is proposed to be an important stabilizing amino acid involved in inter-subunit stability (Fig. 1C). These latter data are consistent with the model that ligand binding results in hexamer formation and that this leads to lowered solvent accessibility and/or structural fluctuations of the “click” motif containing Tyrosine-340.

In domain II, also known as the insertion domain, deuterium suppression induced by ATP includes segments containing amino acids 179–194 and 275–282 (Fig. 6A; and (Supplementary Fig. 5B; in white). This domain is proposed to bind DNA and/or RNA. In domain III, which is also part of the ATP pocket, we were unable to map peptides to almost half of the domain using pepsin digestion (see grey regions in Fig. 6A). However, in the remaining part of domain III, deuteration suppression was observed in segments comprising amino acids 355–371 and 412–433 (Fig. 6A and (Supplementary Fig. 5B; in blue)). Other previously dissected functional domains of Reptin were also examined. The Arg finger (R353) was not covered by peptic peptides. The sensor I region (amino acids M326 - N329) exhibited no changes in deuteration and the sensor II region (amino acids T397 - A402) was also not covered by peptic peptides. An inter-subunit protein-protein interface (amino acids 428–439) were suppressed by ligand (Supplementary Fig. 5C; in white).

Next, we compared the deuteration of Reptin^{Y340A} and Reptin^{Y340A} with 1 mM ATP to determine the effects of reduced oligomerization (Fig. 3B) using deuteration as an assay (Fig. 5 and Fig. 6B). The presence of ATP induced structural changes in Reptin^{Y340A} which are similar to that observed with wild-type Reptin (Figs. 4 and 6; A vs B). Peptic segments that exhibited slower deuterium exchange as a result of ATP binding are highlighted in Fig. 6B (in red). In contrast to wild-type Reptin, there are two segments located in domain I and domain II (amino acids 114–123 and amino acids 275–282 respectively; highlighted in Fig. 6B, in blue colour) that exhibited increase in deuteration after incubation with ATP. These data suggest that this region in Reptin^{Y340A} becomes more exposed to solvent after ATP binding and is suggestive of a lower rate of hexamer formation. Decreases in deuteration were detected similarly as in wild-type Reptin in domain I (segments 51–74 and 90–109) and domain III (segments 355–399 and 411–433).

Interestingly no changes of deuteration were detected in the Y340 peptide docking motif (Y340) in the Reptin^{Y340A} mutant but decreases in deuteration was observed after the amino acid at position 355 (Fig. 5B and Fig. 6B). The neutral effects of ligand on the deuteration of the Y340 peptide docking motif in Reptin^{Y340A} protein are consistent with the model that loss of Tyrosine 340 precludes full “hexamerization” of the Reptin monomer.

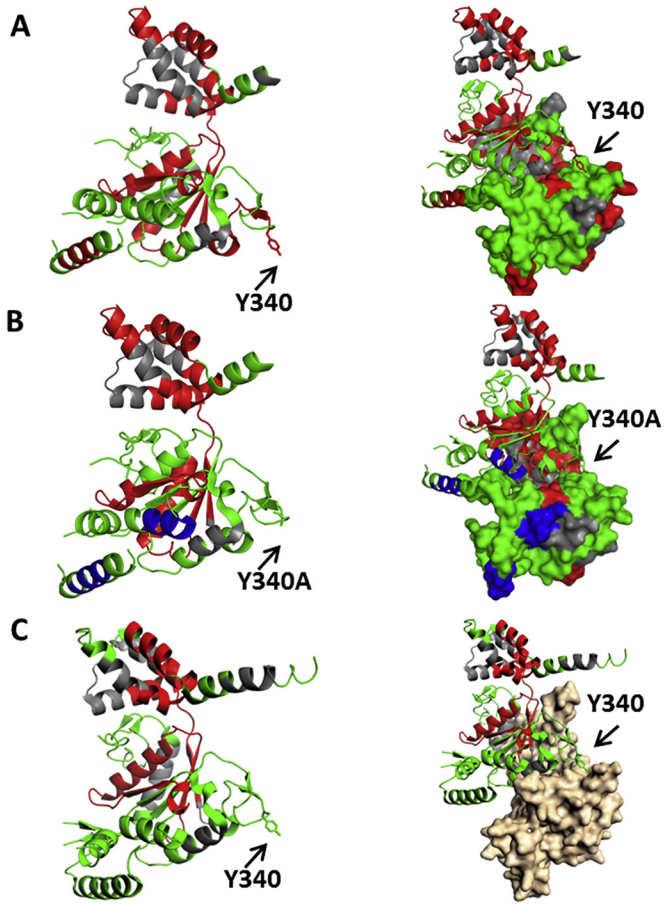


Fig. 6. Localization of deuteration flux superimposed on the three-dimensional structure of Reptin. *(A and B; wild-type Reptin and the Y340A mutant form of Reptin, respectively)*. The crystal structures of the Reptin containing one or two subunits were extracted from the hexamer crystallized as PDB code 3UK6 and the image contains regions highlighted by colors according to changes in deuteration. Red regions indicate areas of decreased deuterium exchange larger than 5% in the presence of ATP and blue regions indicate areas of increased exchange larger than 5%, whilst green indicates no significant change and grey indicates regions with no peptic peptides detected by HDX-MS. (A) The Tyrosine finger (Y340) is indicated by an arrow. (B) The Tyrosine finger (Y340) mutated to alanine is indicted by an arrow. (C). Crystal structures of the monomeric Reptin and the Reptin-Pontin complex (PDB code 2XSZ) with regions highlighted by colors according to changes in deuteration. Red regions indicate areas of decreased exchange larger than 5% in the presence of ATP, while green indicates no significant change and grey indicates regions with no peptides detected by HDX-MS. The Tyrosine finger (Y340) is indicated by arrow. The Pontin is colored light pink. The image was created in PyMOL [23]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Finally, as a control, we compared the deuteration of the Reptin-Pontin complex and Reptin-Pontin complex with 1 mM ATP (Fig. 5 and Fig. 6C). Deuterium incorporation was reduced by ATP in two segments of Reptin (Fig. 5). Peptides protected by ATP in the HDX-MS analysis are highlighted on the structure of Reptin (Fig. 5). The first protected segment with reduced deuteration includes amino acids 63–109 and comprise the Walker A motif. This result is similar to that for wt Reptin and is in accordance with ATP binding to the ATP pocket. The second segment which showed reduced deuterium uptake (amino acids 355–433) belongs to domain III. As seen for Reptin^{Y340A}, the decrease of deuteration caused by ATP was observed in the region following the Y340 peptide docking motif. This suggests that the Y340 peptide docking motif does not participate in the formation of the Reptin-Pontin complex. By contrast, the Reptin-Pontin complex incubated with ligand,

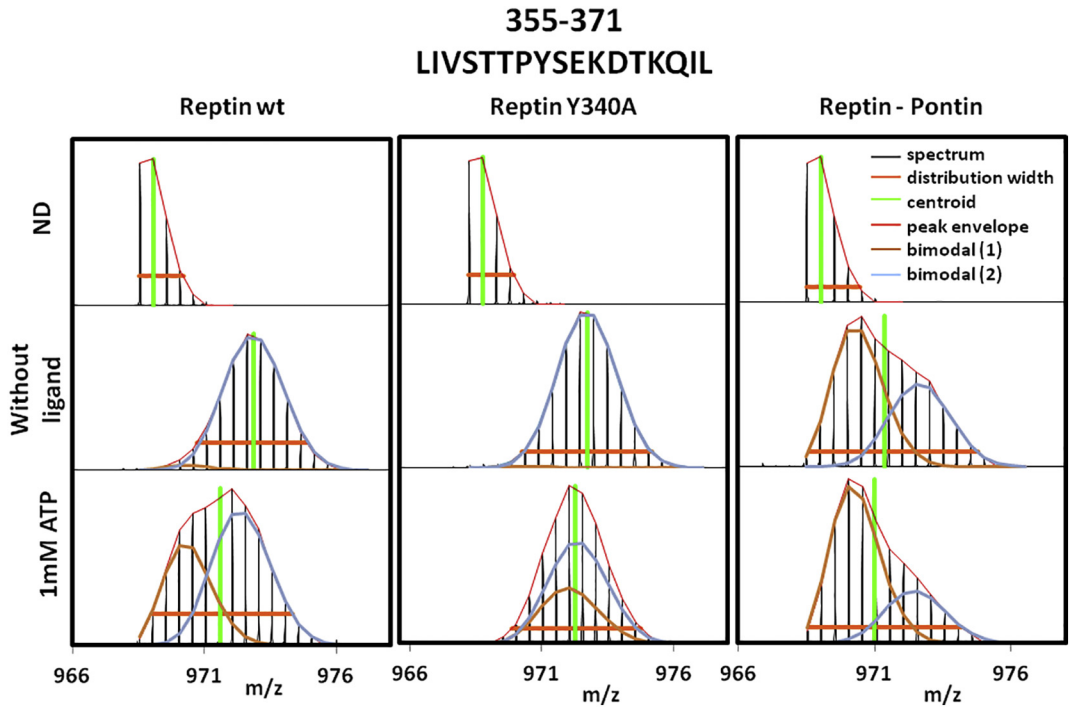


Fig. 7. Bimodal isotopic envelopes. Bimodal isotopic patterns were detected in several peptides. Shown are mass spectra of the peptide containing AAs 355–371. Spectra are shown for all three samples (Reptin wt, Reptin Y340A, and Reptin-Pontin) and for these conditions: undeuterated, sample without ligand and sample with 1 mM ATP. Clear bimodal patterns were detected in Reptin wt after incubation with ATP and in Reptin-Pontin regardless of ATP. On the other hand, Reptin Y340A did not show clear bimodal distributions in any peptides. This is consistent with the data showing that mutation of tyrosine 340 alters Reptin oligomerization. All these findings support our interpretation that bimodal patterns are caused by presence of monomeric/oligomeric state of Reptin rather than EX1 kinetics.

no changes of deuteration were detected in domain II (Fig. 5).

3.5. Bimodal distribution in MS spectra

A summary of Reptin oligomer dynamics is reviewed in Fig. 4. The oligomer dynamics in the presence of ligand can also change as a function of the absolute concentration of Reptin. The concentration of Reptin decreases upon dilution with D_2O. We might predict, therefore, that the HDX methodology can give rise to transient, distinct populations of Reptin as the monomer begins to assemble in the presence of ligand. When analyzing the HDX-MS data, we observed bimodal distributions of several wild-type Reptin peptides in samples with ATP that were not present in the Reptin^{Y340A} mutant. An illustration of bimodal spectra is in Fig. 7. These peptides were located in the ATP pocket, specifically around the walker A motif in domain I and around the Y340 peptide docking motif in domain III. Notably, the bimodal isotopic envelops occurred only in peptides that exhibited conformational changes upon ligand binding. The Reptin-Pontin complex also exhibited bimodal distributions in the presence of ligand (Fig. 7). A bimodal distribution could result from EX1 HDX kinetics in which the rate of conformational change of the corresponding peptide is slower than the HDX rate. The other possible explanation of bimodal distribution is the presence of two different protein isoforms, and here it may arise from the distribution of monomeric and oligomeric forms of Reptin. This is supported by the fact that no bimodal distribution was detected in Reptin^{Y340A} (data not shown) and further, that the change in the pattern of deuteration appeared after the binding of ATP, which is known to initiate Reptin oligomerization. Moreover, the location of peptides with bimodal isotopic patterns also supports this interpretation, because the ATP pocket is located in between two Reptin molecules in the Reptin oligomer.

A complication of these data interpretation is that HDX-MS provides only an average of the deuteration of a given peptide length and does not necessarily provide fine-resolution mapping of a protein-protein interface. For example, our prior study discussed, using AGR2 as a model protein [36], that higher levels of deuterium suppression in a small stretch of the polypeptide are compensated by elevated deuteration in adjacent regions (due to conformational changes) so that the net deuteration in larger peptides is different than the smaller peptide. Thus, it is possible that this methodological limitation impacts on the number of deuterons measured in a given peptic fragment (as in Fig. 5). This limitation did not preclude the identification of regions whose deuteration can be altered through mutation of Y340. In addition, oligomeric proteins present a different problem which might relate to bimodal distributions (Fig. 7). When Reptin assembles into an oligomer, we do not know which monomer a given peptide originates from within the hexamer. A solution to this might be isotopic labelling of a Reptin pool so that upon mixing with normal isotopic Reptin, inter-subunit interactions might be revealed in the mixed olgiomer.

3.6. Use of molecular dynamic (MD) simulations to measure Reptin oligomerization dynamics

We carried out MD simulations to probe the above interpretations (Fig. 4). It is evident that the Y340 sidechain remains embedded in the adjacent subunit (Fig. 8A and C; Fig. 9C), stabilized by a hydrogen bond with the sidechain of Thr333 and also by packing interactions with the sidechains of Arg330 and Arg336 (all 3 residues belonging to the neighboring chain); this is seen in the MD simulations of both wild-type Reptin and in the D299N Reptin mutant (Supplementary Movie 1 and Supplementary Movie 3). In contrast, in the Y340A Reptin mutant (Fig. 8B; Fig. 9D), the sidechain Ala has no polar groups to form the hydrogen bond, resulting in the loop containing Y340 moving away by about 5 Å. The buried surface areas between the monomers (Fig. 9A in the wild-type Reptin and the D299N mutant) remain similar in their liganded states while the *apo* form of wild-type Reptin samples two states – one that resembles the packed liganded states and the other more open state which is similar to that sampled by the Y340A mutant

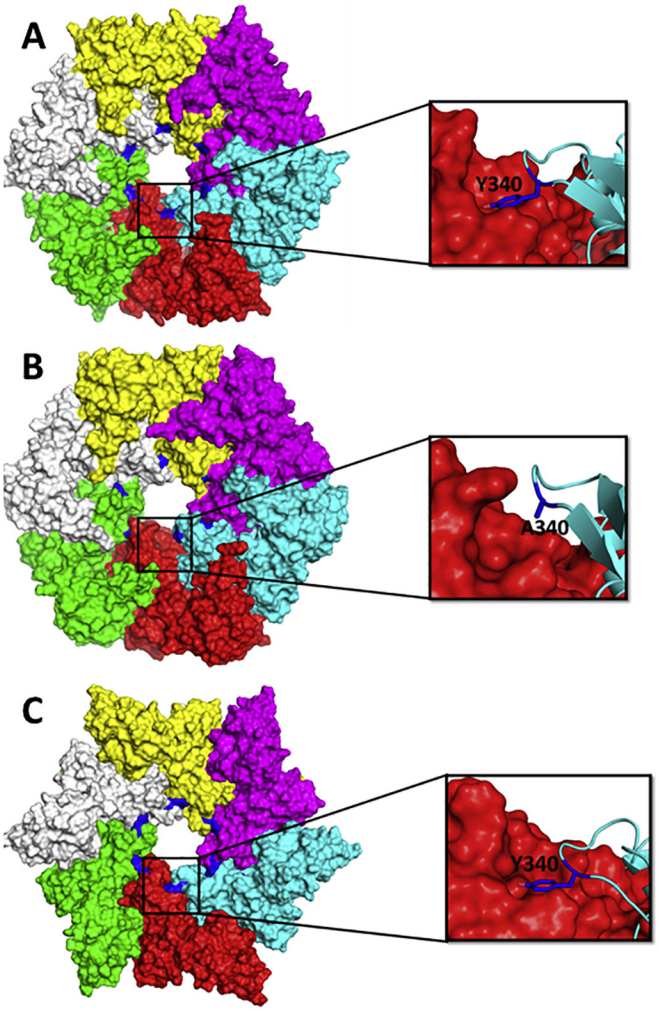


Fig. 8. MD simulation guided modeling of Reptin oligomerization. A snapshot is depicted from the MD simulations of (A). wild-type Reptin, (B). the Y340A mutant form of Reptin and (C). the D299N mutant form of Reptin, as hexamers. Although the Y340A mutant form of reptin may not exist as a hexamer, the structural changes during the MD simulations, which were initiated from the hexameric wild type structure, show that the mutation results in loss of key stabilizing interactions, which would likely result in the dissociation of the hexamer (this dissociation would require very long simulation times). Each monomer unit is colored differently with the Tyrosine finger (Y340) highlighted in blue colour. The data can be visualized in Supplementary movies 1-3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

which loses key interactions that stabilize the hexamer (Supplementary Movie 2).

As a control, D299N Reptin is very stable because the interface near the sidechain of D299 is anionic with several negatively charged groups (arising from the sidechains of E107, E300 and the phosphate of the ligand in the same chain as D299 and the sidechains of D349 and D352 from the neighboring chain, all within 8–9 Å of D299). The loss of the anionic sidechain of D299 in the D299N mutation results in the attenuation of the repulsions; in addition, the N299 sidechain is further stabilized by a water mediated interaction with D349 from the neighboring chain (Fig. 9A and B). This is also apparent in snapshots from the MD simulations (Supplementary Movies 1-3).

4. Discussion

Reptin forms a model AAA+ superfamily member with which to develop new approaches for dissecting stages in ligand-mediated oligomer assembly. In the past thirty years, the dominant tools used by proteomics approaches to discovery protein-interactions have been; (i) the yeast two hybrid; and (ii) immuno-precipitations. It is also now known that the vast diversity of protein interaction dynamics in higher eukaryotes appear to be regulated by small linear peptide motifs (27). Linear motifs in peptides can form 'hotspots' that play dominant role in driving protein-protein interactions and might form key target for drugging protein-protein interactions [38,39]. A classic example is the oncogenic protein MDM2; which can form a high affinity interaction with a small synthetic peptide derived from the p53 tumor suppressor [40] and which has been the basis for drug that target the linear peptide binding pocket of MDM2 [41].

Thus, a methodological approach for identifying hotspot protein-peptide interaction sites is using either synthetic peptides or phage-combinatorial peptide libraries [42]. Overlapping synthetic peptide scan assays using peptides derived from the open reading frame of Reptin were used to determine whether self-peptides that interact stably with Reptin itself, could be identified. This approach revealed a relatively high affinity, self-assembly interface that stabilizes the Reptin hexamer (dominating by Tyrosine 340). Only one self-binding motif for Reptin was identified (peptide 28; Fig. 1). For example, peptides derived from the an inter-subunit interface region of Reptin (amino acids ~420–440) did not bind Reptin in this assay. This highlights a limitation of this assay; it can not necessarily identify any protein-protein contact dominated by linear peptide motifs. However, peptides that do bind in such assays form compelling models to dissect out mechanisms of linear-peptide-protein docking.

The single dominating self-peptide maps to the Reptin hexameric "rim" (Fig. 1B), with a Tyrosine residue (Y340) forming an inter-subunit stabilizing anchor (Figs. 8-9). The subunit that interacts with the tyrosine-containing donor peptide forms a small acceptor pocket (Fig. 9). Size exclusion chromatography affirmed that ligand-induced oligomerization of mutant Y340A Reptin is attenuated (Fig. 3B).

There is presumably combinatorial diversity in this Y340 donor-acceptor docking reaction. Although the Y340 side chain forms an important stabilizing contact in a Reptin-Reptin homo-oligomer, it might not be as important in a Reptin-Pontin hetero-oligomer (Fig. 6C). Further, whether the 'rim' of a Pontin-Pontin oligomer also forms a hotspot stabilizing contact in this vicinity requires additional functional studies. An unexpected result from our studies, that complements the identification of Y340 as an important driver of Reptin assembly, was the impact of the D299N mutation on oligomerization.

The D299N version of Reptin, lacks ATPase activity and has classically been used as a 'loss-of-function' mutant [33]. However, this mutant has an elevated ATP-binding activity [13]. In our current study, we can interpret the data that the D299N mutant produces a spontaneous oligomer in the absence of ligand (Fig. 3C). Molecular Dynamics simulations show that the ability of Y340A Reptin to form stable oligomers is attenuated because of the loss of the hydrogen bond and packing afforded by the sidechain of Y340 against neighboring monomers (Fig. 8-9). However, in the D299N Reptin mutant, destabilizing, negatively charged inter-subunit contacts are removed, resulting in the stabilization of the hexamer in the absence of ligand (Fig. 9; Movie 3). This gives a more complex interpretation to cell based studies using the D299N mutant of Reptin. Its inhibitory effect on growth could be because of (i) defects in ATPase function; (ii) elevated ATP-binding; or (iii) elevated spontaneous oligomerization (this study). In the latter case, since Reptin's specific activity in protein-binding assays is regulated by its monomer and oligomerization status (Fig. 4), then in cells the Reptin^{D299N} mutant could be inhibitory for cell growth because of alterations in its interactome.

In conclusion, our pipeline approach for dissecting 'hotspot' interactions in one AAA+ family member provides a blueprint to produce functional mutants based on self-assembly peptide motif scans. Our data highlight a novel linear motif in Reptin that forms an inter-subunit

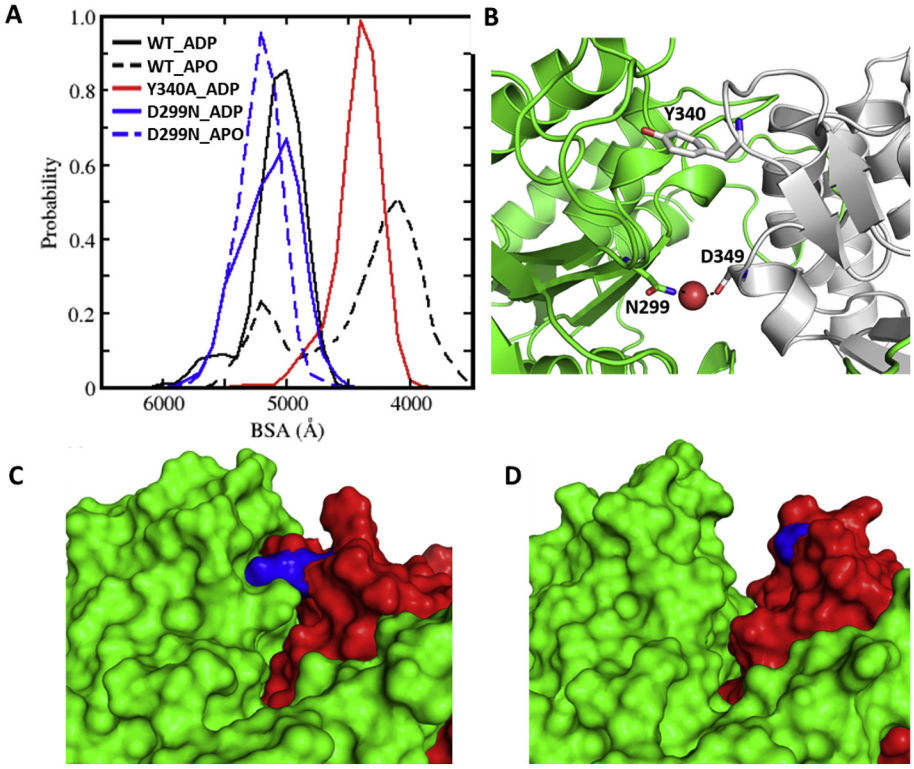


Fig. 9. Evaluation of the D299N and Y340A forms of Reptin. (A). Distribution of buried surface area (BSA) of the inter-subunit protein-protein interface calculated from the conformations sampled during the MD simulations of wildtype and mutant Reptin hexamers in their *apo* and ADP bound states. The BSA was calculated by subtracting the solvent accessible surface area (SASA) of the inter-subunit protein-protein interface from the sum of the SASA of each monomer. (B). MD snapshot of the Reptin hexamer showing water mediated interactions between two monomers. Residues (N299 from monomer A and D349 from monomer B) involved in water mediated interactions are highlighted in sticks, the water molecule is shown as a red sphere and water-mediated hydrogen bond interactions are shown as dashed lines. (C). Snapshot from the MD simulations highlighting the tight packing between the two monomers (colored green and red) in the case of wild-type Reptin and (D). loss in packing interactions due to the Y340A mutation. Reptin is depicted as surface, and Y340 and A340 are highlighted in blue colour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

bridge that appears to play a role in hexamer stability. Such mutant forms of Reptin (Y340A) can be used in the future in cell systems to dissect the interactome of Reptin as a function of monomer to oligomer states (Fig. 4). There is a precedent for the ability of small peptides to interact with dynamic forms of Reptin; a recent biochemical study evaluated the impact of histones on Reptin and Pontin transcriptional regulation [43]. The authors identified histone-3 peptides that not only regulate the monomer-oligomer transition of Reptin and Pontin but showed that endogenous proteins immuno precipitated by monomeric Reptin are distinct from proteins that interact with oligomers. These data highlight the functional importance of Reptin oligomerization in controlling its interactome (as in Fig. 4 and [43]).

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jprot.2019.02.012.

Acknowledgements

The work was supported by: the Czech Science Foundation 16-20860S (PM, LH) and 16-07321S (BV, TH), the project MEYS – NPS I – LO1413, and MH CZ - DRO (MMCI, 00209805); the BBSRC RASOR consortium (BB/C511599/1; United Kingdom); Cancer Research UK (C21383/A6950); The International Centre for Cancer Vaccine Science project carried out within the International Research Agendas programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund; A*STAR, Singapore and NSCC, Singapore.

References

[1] R.T. Sauer, D.N. Bolon, B.M. Burton, R.E. Burton, J.M. Flynn, R.A. Grant, G.L. Hersch, S.A. Joshi, J.A. Kenniston, I. Levchenko, S.B. Neher, E.S. Oakes, S.M. Siddiqui, D.A. Wah, T.A. Baker, Sculpting the proteome with AAA(+) proteases and disassembly machines, Cell 119 (1) (2004) 9–18.
[2] R.T. Sauer, T.A. Baker, AAA+ proteases: ATP-fueled machines of protein destruction, Annu. Rev. Biochem. 80 (2011) 587–612.
[3] P.M. Matias, S. Gorynia, P. Donner, M.A. Carrondo, Crystal structure of the human AAA+ protein RuvBL1, J. Biol. Chem. 281 (50) (2006) 38918–38929.
[4] S. Gorynia, T.M. Bandeiras, F.G. Pinho, C.E. McVey, C. Vonrhein, A. Round, D.I. Svergun, P. Donner, P.M. Matias, M.A. Carrondo, Structural and functional insights into a dodecameric molecular machine - the RuvBL1/RuvBL2 complex, J. Struct. Biol. 176 (3) (2011) 279–291.
[5] M. Petukhov, A. Dagkessamanskaja, M. Bommer, T. Barrett, I. Tsaneva, A. Yakimov, R. Queval, A. Shvetsov, M. Khodorkovskiy, E. Kas, M. Grigoriev, Large-scale conformational flexibility determines the properties of AAA+ TIP49 ATPases, Structure 20 (8) (2012) 1321–1331.
[6] J. Jin, Y. Cai, T. Yao, A.J. Gottschalk, L. Florens, S.K. Swanson, J.L. Gutierrez, M.K. Coleman, J.L. Workman, A. Mushegian, M.P. Washburn, R.C. Conaway, J.W. Conaway, A mammalian chromatin remodeling complex with similarities to the yeast INO80 complex, J. Biol. Chem. 280 (50) (2005) 41207–41212.
[7] T.L. Clarke, M.P. Sanchez-Bailon, K. Chiang, J.J. Reynolds, J. Herrero-Ruiz, T.M. Bandeiras, P.M. Matias, S.L. Maslen, J.M. Skehel, G.S. Stewart, C.C. Davies, PRMT5-Dependent Methylation of the TIP60 Coactivator RUVBL1 Is a Key Regulator of Homologous Recombination, Mol. Cell. 65 (5) (2017) 900–916 e7.
[8] L. Gnatovskiy, P. Mita, D.E. Levy, The human RVB complex is required for efficient transcription of type I interferon-stimulated genes, Mol. Cell. Biol. 33 (19) (2013) 3817–3825.
[9] S.G. Cho, A. Bhoumik, L. Broday, V. Ivanov, B. Rosenstein, Z. Ronai, TIP49b, a regulator of activating transcription factor 2 response to stress and DNA damage, Mol. Cell. Biol. 21 (24) (2001) 8398–8413.
[10] P. Bellosta, T. Hulf, S. Balla Diop, F. Usseglio, J. Pradel, D. Aragnol, P. Gallant, Myc interacts genetically with Tip48/Reptin and Tip49/Pontin to control growth and proliferation during Drosophila development, Proc. Natl. Acad. Sci. U. S. A. 102 (33) (2005) 11799–11804.
[11] S. Rashid, I. Pilecka, A. Torun, M. Olchowik, B. Bielinska, M. Miaczynska, Endosomal adaptor proteins APPL1 and APPL2 are novel activators of beta-catenin/TCF-mediated transcription, J. Biol. Chem. 284 (27) (2009) 18115–18128.
[12] C. Xie, W. Wang, F. Yang, M. Wu, Y. Mei, RUVBL2 is a novel repressor of ARF transcription, FEBS Lett. 586 (4) (2012) 435–441.
[13] M.M. Maslon, R. Hrstka, B. Vojtesek, T.R. Hupp, A divergent substrate-binding loop within the pro-oncogenic protein anterior gradient-2 forms a docking site for Reptin, J. Mol. Biol. 404 (3) (2010) 418–438.
[14] N. Izumi, A. Yamashita, A. Iwamatsu, R. Kurata, H. Nakamura, B. Saari, H. Hirano, P. Anderson, S. Ohno, AAA+ proteins RUVBL1 and RUVBL2 coordinate PIKK activity and function in nonsense-mediated mRNA decay, Sci. Signal. 3 (116) (2010) ra27.
[15] A.S. Venteicher, Z. Meng, P.J. Mason, T.D. Veenstra, S.E. Artandi, Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly, Cell 132 (6) (2008) 945–957.
[16] A. Rivera-Calzada, M. Pal, H. Munoz-Hernandez, J.R. Luque-Ortega, D. Gil-Carton, G. Degliesposti, J.M. Skehel, C. Prodromou, L.H. Pearl, O. Llorca, The Structure of the R2TP Complex Defines a Platform for Recruiting Diverse Client Proteins to the HSP90 Molecular Chaperone System, Structure 25 (7) (2017) 1145–1152 e4.
[17] J. Rosenbaum, S.H. Baek, A. Dutta, W.A. Houry, O. Huber, T.R. Hupp, P.M. Matias, The emergence of the conserved AAA+ ATPases Pontin and Reptin on the signaling landscape, Sci. Signal. 6 (266) (2013) mr1.

[18] A.R. Healy, D.R. Houston, L. Remnant, A.S. Huart, V. Brychtova, M.M. Maslon, O. Meers, P. Muller, A. Krejci, E.A. Blackburn, B. Vojtesek, L. Hernychova, M.D. Walkinshaw, N.J. Westwood, T.R. Hupp, Discovery of a novel ligand that modulates the protein-protein interactions of the AAA+ superfamily oncoprotein reptin, Chem. Sci. 6 (5) (2015) 3109–3116.
[19] J.K. Murray, S.H. Gellman, Targeting protein-protein interactions: lessons from p53/MDM2, Biopolymers 88 (5) (2007) 657–686.
[20] M. Durech, F. Trcka, P. Man, E.A. Blackburn, L. Hernychova, P. Dvorakova, D. Coufalova, D. Kavan, B. Vojtesek, P. Muller, Novel Entropically driven conformation-specific interactions with Tomm34 protein modulate Hsp70 protein folding and ATPase activities, Mol. Cell. Proteomics 15 (5) (2016) 1710–1727.
[21] D. Kavan, P. Man, MSTools—web based application for visualization and presentation of HXMS data, Int. J. Mass Spectrom. 302 (2011) 53–58.
[22] M. Guttman, D.D. Weis, J.R. Engen, K.K. Lee, Analysis of overlapped and noisy hydrogen/deuterium exchange mass spectra, J. Am. Soc. Mass Spectrom. 24 (12) (2013) 1906–1912.
[23] W.L. DeLano, PyMOL, 700 DeLano Scientific, San Carlos, CA, 2002.
[24] D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C.L. Simmerling, W. Smith, J. Swails, J. Wang Walker, R.M. Wolf, X. Wu, P.A. Kollman, Amber14, University of California, San Francisco, 2014.
[25] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, J. Chem. Theory Comput. 11 (8) (2015) 3696–3713.
[26] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (9) (2004) 1157–1174.
[27] W.L. Jorgensen, J. Chandrasekhar, J. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.
[28] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an Nlog (N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.
[29] S. Miyamoto, P.A. Kollman, Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models, J. Comput. Chem. 13 (1992) 952–962.
[30] Y. Bromberg, G. Yachdav, Y. Ofran, R. Schneider, B. Rost, New in protein structure and function annotation: hotspots, single nucleotide polymorphisms and the 'Deep Web', Curr. Opin. Drug Discov. Devel. 12 (3) (2009) 408–419.
[31] P. Tompa, N.E. Davey, T.J. Gibson, M.M. Babu, A million peptide motifs for the molecular biologist, Mol. Cell 55 (2) (2014) 161–169.
[32] H. Benbahouche Nel, I. Iliopoulos, I. Torok, J. Marhold, J. Henri, A.V. Kajava, R. Farkas, T. Kempf, M. Schnolzer, P. Meyer, I. Kiss, E. Bertrand, B.M. Mechler, B. Pradet-Balade, Drosophila Spag is the homolog of RNA polymerase II-associated protein 3 (RPAP3) and recruits the heat shock proteins 70 and 90 (Hsp70 and Hsp90) during the assembly of cellular machineries, J. Biol. Chem. 289 (9) (2014) 6236–6247.
[33] A. Grigoletto, V. Neaud, N. Allain-Courtois, P. Lestienne, J. Rosenbaum, The ATPase activity of reptin is required for its effects on tumor cell growth and viability in hepatocellular carcinoma, Mol. Cancer Res. 11 (2) (2013) 133–139.
[34] L. Hernychova, P. Man, C. Verma, J. Nicholson, C.A. Sharma, E. Ruckova, J.Y. Teo, K. Ball, B. Vojtesek, T.R. Hupp, Identification of a second Nutlin-3 responsive interaction site in the N-terminal domain of MDM2 using hydrogen/deuterium exchange mass spectrometry, Proteomics 13 (16) (2013) 2512–2525.
[35] V. Narayan, V. Landre, J. Ning, L. Hernychova, P. Muller, C. Verma, M.D. Walkinshaw, E.A. Blackburn, K.L. Ball, Protein-protein interactions modulate the docking-dependent E3-ubiquitin ligase activity of Carboxy-terminus of Hsc70-interacting protein (CHIP), Mol. Cell. Proteomics 14 (11) (2015) 2973–2987.
[36] M.A. Mohtar, L. Hernychova, J.R. O'Neill, M.L. Lawrence, E. Murray, B. Vojtesek, T.R. Hupp, The sequence-specific peptide-binding activity of the protein Sulfide Isomerase AGR2 directs its stable binding to the oncogenic receptor EpCAM, Mol. Cell. Proteomics 17 (4) (2018) 737–763.
[37] F. Trcka, M. Durech, P. Man, L. Hernychova, P. Muller, B. Vojtesek, The assembly and intermolecular properties of the Hsp70-Tomm34-Hsp90 molecular chaperone complex, J. Biol. Chem. 289 (14) (2014) 9887–9901.
[38] H. Jubb, A.P. Higueruelo, A. Winter, T.L. Blundell, Structural biology and drug discovery for protein-protein interactions, Trends Pharmacol. Sci. 33 (5) (2012) 241–248.
[39] A. Ciulli, G. Williams, A.G. Smith, T.L. Blundell, C. Abell, Probing hot spots at protein-ligand binding sites: a fragment-based approach using biophysical methods, J. Med. Chem. 49 (16) (2006) 4992–5000.
[40] S.M. Picksley, B. Vojtesek, A. Sparks, D.P. Lane, Immunochemical analysis of the interaction of p53 with MDM2;–fine mapping of the MDM2 binding site on p53 using synthetic peptides, Oncogene 9 (9) (1994) 2523–2529.
[41] L.T. Vassilev, B.T. Vu, B. Graves, D. Carvajal, F. Podlaski, Z. Filipovic, N. Kong, U. Kammlott, C. Lukacs, C. Klein, N. Fotouhi, E.A. Liu, In vivo activation of the p53 pathway by small-molecule antagonists of MDM2, Science 303 (5659) (2004) 844–848.
[42] H. Shimizu, L.R. Burch, A.J. Smith, D. Dornan, M. Wallace, K.L. Ball, T.R. Hupp, The conformationally flexible S9-S10 linker region in the core domain of p53 contains a novel MDM2 binding site whose mutation increases ubiquitination of p53 in vivo, J. Biol. Chem. 277 (32) (2002) 28446–28458.
[43] R. Queval, C. Papin, M. Dalvai, K. Bystricky, O. Humbert, Reptin and Pontin oligomerization and activity are modulated through histone H3 N-terminal tail interaction, J. Biol. Chem. 289 (49) (2014) 33999–34012.